Research Product 2005-09

A Training Technology Evaluation Tool

Stephen C. Livingston
Northrop Grumman Mission Systems

Jean L. Dyer
Army Research Institute

Diadra Swinson
Northrop Grumman Mission Systems

September 2005

Infantry Forces Research Unit

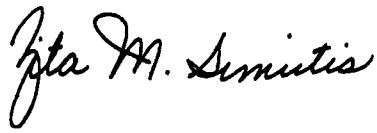
United States Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the Department of the Army Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS
Technical Director



ZITA M. SIMUTIS
Director

Research accomplished under contract
for the Department of the Army

Northrop Grumman Mission Systems

Technical Review by

Michael J. Singer, U.S. Army Research Institute
Jennifer Sommers Tucker, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Product has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Product are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) September 2005	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) July 2003 to June 2004
4. TITLE AND SUBTITLE A Training Technology Evaluation Tool		5a. CONTRACT OR GRANT NUMBER DASW01-99-D-0013 (DO#30)
		5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S Stephen C. Livingston (Northrop Grumman Mission Systems), Jean L. Dyer (U.S. Army Research Institute), and Diadra Swinson (Northrop Grumman Mission Systems)		5c. PROJECT NUMBER A790
		5d. TASK NUMBER
		5e. WORK UNIT NUMBER 215
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Northrop Grumman Mission Systems 3565 Macon Road Columbus, Georgia 31907		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IJ 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Research Product 2005-09

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter Expert: Jean Dyer

14. ABSTRACT *(Maximum 200 words)*: A Training Technology Evaluation Tool was developed to help procurers and developers of training technologies to make informed decisions and to improve the overall effectiveness of training technologies. The tool provides estimates of training technology effectiveness, based on expert ratings. The overall ratings reflect the initial performance of the Soldiers to be trained on the tasks covered by the training technology, task and subtask characteristics, learning difficulty, residual performance deficits after using the technology, physical and functional similarities to the operational environment/equipment, and training transfer. The questions in the tool are presented via an Excel program, whereby all mathematical calculations are automated and transparent to the user. The tool can be used at any stage of training technology design and implementation.

15. SUBJECT TERMS

Training Training Technology Training Technology Evaluation Training Device
Training assessment Simulator

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	85	Ellen Kinzer Technical Publication Specialist (703) 602-8047

Research Product 2005-09

A TRAINING TECHNOLOGY EVALUATION TOOL

Stephen C. Livingston
Northrop Grumman Mission Systems

Jean L. Dyer
Army Research Institute

Diadra Swinson
Northrop Grumman Mission Systems

Infantry Forces Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, VA 22202-3926

September 2005

Army Project Number
622785A790

Personnel Performance
and Training Technology

A TRAINING TECHNOLOGY EVALUATION TOOL

EXECUTIVE SUMMARY

Research Requirement:

With the fielding of complex operational systems, being able to determine what is or could be effective training technologies for these systems is crucial. Since the 1960s, the Army has examined various methodologies to determine the efficiency and effectiveness of training aids, devices, simulators, and simulations (TADSS) technologies. This project was an effort to develop an evaluation tool to examine the feasibility and potential value of existing and emerging training technologies, to include those for the Future Force, and a tool that addressed weaknesses in prior approaches. The requirement was to enable assessments of TADSS and interactive multi-media instruction via a user-friendly automated tool, useable and modifiable by a wide-audience to include researchers, military training developers, and military subject matter expert (SMEs).

Procedure:

The project started with an assessment worksheet provided by the government. However, it was determined that this worksheet was not sensitive to differences in training technologies. Further investigation led to revising a procedure developed in the 1980s called the Device Effectiveness Forecasting Technology (DEFT). The DEFT was enhanced, revised to evaluate more types of training, automated using common software tools, and adapted to be user-friendly and understandable. A scoring guide was developed to help SMEs better understand the evaluation tool results. Evaluations of the Tool using SMEs and four different training technologies, including simulator and non-simulator technologies, were conducted.

Findings:

The Training Technology Evaluation Tool provided three levels of analyses by which SMEs could evaluate training technologies at different stages of development. Interviewing the SMEs after they completed their ratings provided insights regarding training priorities and assumptions. Some differences in ratings were found when different SMEs assessed the same technology using the same level of analysis. In general, ratings were consistent across the three levels of analysis when the same SME did all the assessments. The highest level of analysis required the most knowledge about the training domain and the training technology. Further assessments on a broader scale are required to fully determine the efficiency and effectiveness of the Tool.

Utilization and Dissemination of Findings:

The Training Technology Evaluation Tool derived from this work is easy to use. It allows military SMEs and trainers to evaluate both emerging and legacy training technologies to help determine the efficient and effective ways to train the tasks and subtasks that Soldiers require. Also, it can be used to evaluate future design plans and prototypes of training technologies to determine their potential value to Soldiers and the Army, and how they could be improved. Use of the Training Technology Evaluation Tool during TADSS development and design could save both time and money. Such use would help ensure efficient and effective devices, simulators, and simulations are produced from inception and reduce the time required to field them. Army and TRADOC program managers can use the Tool throughout the procurement cycle for TADSS devices that support new equipment and doctrine.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

A TRAINING TECHNOLOGY EVALUATION TOOL

Introduction

Since the early 1960s, the Army has searched to find a methodology for determining the efficiency and effectiveness of training aids, devices, simulators, and simulations (TADSS) technologies. With the advent of systems such as the Abrams tank, the Bradley Fighting Vehicle, the Apache Longbow, the Paladin Howitzer, and highly complex communications equipment, the need for effective, efficient training has become even more critical. The Army has invested heavily in digital communications technologies to better enable combat forces to see and control the fight, further complicating the training challenge. Due to these rapidly developing and increasingly complex systems, the search for efficient and effective training technologies has become even more critical. To further compound the situation, the demand for resources has increased exponentially, causing the Army to consider existing commercial-off-the-shelf items as solutions to training problems. Since training resources are limited, decisions regarding the best training technologies are critical. It is therefore incumbent upon the Army to find a methodology to help predict the most efficient and effective technologies for training. This research project is part of that effort.

Several Army Research Institute (ARI) research efforts have been dedicated to finding a method to assess the efficiency and effectiveness of Army TADSS and training technologies. These efforts include Wheaton, Fingerman, Rose, and Leonard (1976), Narva (1979), Tufano and Evans (1982), Rose, Wheaton, and Yates (1985a, 1985b), and Sticha, Gibbons, and Singer (1993). The tool presented in this report is based on these prior efforts, but expanded the scope of technologies that can be evaluated, and was converted to a user-friendly format.

The report describes the effort to design and develop a systematic procedure, a Training Technology Evaluation Tool, to help assess the potential value of emerging and current training technologies. Included in the report is a description of the research performed along similar lines in the past; how that research impacted the evolution of the Tool under the auspices of this effort; and the results of the research performed during the development of the Tool.

Objectives

The general requirements of a Training Technology Evaluation Tool, which provided the framework for tool development, were:

- It should evaluate emerging and legacy technologies.
- It should be able to evaluate simulations, emulations, and interactive multimedia instruction (IMI).
- It should be usable by subject matter experts, training developers, and behavioral scientists.
- It should be automated, but at the same time be simple to use and simple to modify, should the need arise.

- It should either eliminate or reduce biases an evaluator or user might have towards any particular technology.
- It should have a scoring guide to help trainers use the results.

Prior Training Device Assessment Tools and Procedures

Training Assessment Worksheet

A training assessment worksheet provided by the government as a starting point touched on many of the factors that must be addressed to determine what training technologies will help address training requirements and needs. A sample of the training assessment worksheet in its original form is found at Appendix B.

However, the questions in the training assessment worksheet appeared to be too general to generate a quantifiable evaluation. The worksheet was revised in an attempt to produce a Tool that would yield a quantifiable measure of training effectiveness. A sample of the revised worksheet is at Appendix C. Although a scale was added for several questions, the preponderance of questions required "yes" and "no" responses. The combination of "yes," "no," and scaled questions still did not yield sufficient differentiation to discern between a "good" and "bad" training technology. At that point, an additional literature review was conducted to find a means of recording and scoring information about training technologies that would provide greater, yet appropriate discrimination.

Training Device Effectiveness (TRAINVICE)

Wheaton, et al. (1976) investigated techniques that could evaluate training devices and predict their effectiveness. The models developed from this and related research became known as TRAINVICE. The research relied on the capability to measure the transfer of skills from the training technology to the operational equipment or situation. The TRAINVICE approach assumed that transfer of skills could be measured in terms of three components: transfer potential, learning deficit, and training techniques. In turn, these components would forecast the effectiveness of a training device or technology through the analyses in five areas called:

- Task Commonality Analysis
- Physical Similarity Analysis
- Functional Similarity Analysis
- Learning Deficit Analysis
- Training Techniques Analysis

In all areas the training device or technology was compared to the operational equipment or situation presented in training. Tufano and Evans (1982) referred to the system devised by Wheaton, et al. (1976) as TRAINVICE-A, an attempt to determine predictive effectiveness of a training device. The task commonality analysis compared the operational subtasks of the actual equipment or situation with the training subtasks covered by the training technology. For each operational subtask not covered in the training technology a rating of "0" was given. For each operational subtask covered by the training technology a rating of "1" was given. Additional knowledge and skills were evaluated in a similar manner.

Two problems with the TRAINVICE methodology were identified. First, one of the goals established for the Training Technology Evaluation Tool was to allow subject matter experts (SMEs) to perform the evaluation. Yet the complexity of determining subtasks, skills, and knowledge, as defined in TRAINVICE, appeared to be difficult to achieve with any degree of confidence by SMEs. The second problem was that tasks and subtasks for operational equipment or situations were not always clearly defined for the training technologies.

This appraisal of TRAINVICE was supported during interviews with SMEs for each of the training technologies evaluated during this research project. One of the technologies evaluated was a marksmanship training device, which goes well beyond the requirements for basic and advanced rifle marksmanship. It is capable of providing basic and advanced rifle marksmanship training to an infantry squad. It also provides training for Soldiers as to when and when not to engage certain targets in scenario-driven exercises, and marksmanship training for infantry squads under simulated combat conditions. However, the device configuration fielded at the time did not take into account the requirement for firing the M4 carbine using specific sights such as the thermal weapon sight (TWS) or the close combat optic (CCO). Since all the tasks for firing with advanced optics are capable of being trained in the real world situation but not with the marksmanship trainer, using Wheaton's et al. (1976) formulae to achieve a task commonality analysis, the device would have scored poorly. Of eighteen operational subtasks required for advanced rifle marksmanship, the device only provides training for seven. Therefore eleven subtasks would have received a rating of zero.

Another problem was the dichotomous scales. Using TRAINVICE-A's binary scoring system of "1" and "0" did not provide sufficient discrimination between important operational subtasks and "nice to have" operational subtasks.

It became apparent that the task commonality analysis in TRAINVICE-A would either have to be completely revised or be eliminated. However, the other aspects of TRAINVICE-A did seem to have great merit. The Physical and Functional Similarity analyses implied a relationship to realism, while the Learning Deficit analysis seemed to answer the aspects concerning the need for a training technology. The process of conducting a Training Techniques Analysis provided only a marginal scoring differentiation. According to Wheaton et al. (1976), the Training Techniques Analysis did not always give credit to the technology being evaluated for having some good instructional features.

The Physical Similarities Analysis (Wheaton et al., 1976) was derived by a comparative analysis of the training technology to the operational equipment or situation. It simply reflected how much the training technology and operational equipment or situation "looked alike." For example, if the training technology is a driving simulator for an M1 tank driver, are controls present such as the throttle on the steering "t-bar", and are displays such as a speedometer and a vision block to the terrain outside present? Function was not considered when performing the physical similarities analysis. The more the training technology "looked" like the operational equipment or situation, the higher the score. However, the final Physical Similarities Analysis score was reduced to a dichotomous "1" or "0" score.

The Functional Similarities Analysis was based on the manner in which the operator assimilates information and processes it. Using the same analogy of the driver's station for an M1 tank, the function of the accelerator or throttle is assessed, as well as what information the driver uses to determine speed. If both the speedometer and the terrain visible in the vision block move according to whether the driver is accelerating or decelerating, then the functional analysis is identical to that in the operational equipment, the M1 tank. Like the physical analysis, the more realistic the functions in the training technology, the higher the raw score. In the end, Functional Similarities Analysis was reduced to a score of either "1" or "0."

The Learning Deficit from TRAINVICE-A compared what Tufano and Evans (1982) referred to as a repertory scale to a criterion scale. The repertory scale required an assessment of the degree of each individual skill and knowledge that a Soldier already has before training. The criterion scale required an assessment of what each individual skill and knowledge the Soldier must attain to achieve proficiency. However, the ability of an Army SME to perform these evaluations on a quantifiable scale was not present within the TRAINVICE-A methodology.

The TRAINVICE study by Tufano and Evans (1982) indicated a need for further investigation concerning what they referred to as TRAINVICE C, the same as what Narva (1979) referred to as TRAINVICE II. Narva (1979) raised the question of "why" each particular task or subtask should be included in the technology. He proposed that an assessment be performed to determine the Training Criticality of each task and subtask, and then assess the Training Difficulty for each task or subtask. Training Difficulty would be a judgment of the degree of difficulty that would be expected in training to reach the desired level of proficiency. Training Difficulty, although a judgment, would have to be quantified and measured for each activity covered by the training technology.

Device Effectiveness Forecasting Technique (DEFT)

The Device Effectiveness Forecasting Technique (DEFT) by Rose et al. (1985a, 1985b) was an enhancement to the TRAINVICE-A instrument. The most significant modification made was the recognition that not only is the analysis of potential transfer important, but the learning deficits both before and after training with the technology were equally important.

It also resolved the problems with TRAINVICE that were cited previously. The DEFT quantified the extent to which a Soldier could meet an expected standard of performance using the training device (Rose et al., 1985a, 1985b). Rose et al. (1985b) provided a solution to the problem raised by Narva (1979) regarding the need to assess the difficulty of each task and subtask to be trained. The DEFT included questions on deficit skills and knowledge, what training was needed to bring the Soldier to required performance levels, and the difficulty involved in being able to train the Soldier to those standards. By stating the questions in that form, they were easier for SMEs to evaluate. The standardization of the questions and the answers permitted automation. This approach was consistent with our initial assessment that SMEs and instructors should complete the evaluation of a training technology, as they are in the best position to assess where difficulties would be encountered with the "average" Soldier to be trained.

Another important modification from TRAINVICE-A was that the DEFT incorporated three levels of analysis. Level I was the least detailed analysis based on the least amount of information, while Level III provided the most detailed analysis based on detailed information and experience concerning the training technology and the system tasks.

DEFT required an analysis of four areas called:

- Training Problem,
- Acquisition Efficiency,
- Transfer Problem, and
- Transfer Efficiency.

The evaluator selects the tasks or subtasks to be considered. Therefore, the evaluation can be conducted as granularly as necessary.

With DEFT, the Training Problem (*TP*) analysis defined the deficiency in skills and knowledge that a Soldier can have with regard to the criterion performance (Performance Deficit, *PD*), and the difficulty Soldiers would have in overcoming the deficiencies (Learning Difficulty, *D*). An assessment of the quality of training provided by the technology, in terms of instructional features and principles found in the training technology that help Soldiers overcome their deficiencies, was called the Acquisition Efficiency (*AE*) analysis. The first two analyses (TP and AE) resulted in a total Acquisition (*A*) score.

The third analysis, called the Transfer Problem (*TRP*) analysis, determined the Residual Performance Deficit (*RPD*) after the Soldier had been trained using the training technology. It also determined the difficulty involved in overcoming that deficit, called Residual Learning Difficulty (*RLD*). The *TRP* analysis also considered Additional Deficits (AD) created by Physical Dissimilarities (*PS*) and Functional Dissimilarities (*FS*) between the training technology and the operational equipment or situation.

The fourth analysis provided an index of transfer efficiency (*TT*), reflecting how well the training technology will promote the transfer of learning from the technology to the actual equipment or situation. The resultant scores from the transfer efficiency (*TT*) analysis and the transfer problem analysis (*TRP)* determine transfer (*T*).

Total Effectiveness (Σ) was the index that represented the overall training technology effectiveness. It was obtained by combining the Acquisition (*A*) score and the Transfer (*T*) score. The lower the Total Effectiveness value (Σ), the more effective the training technology. Figure 1 illustrates the relationships among the analysis components for the DEFT. These same components were used in the Training Technology Evaluation Tool.

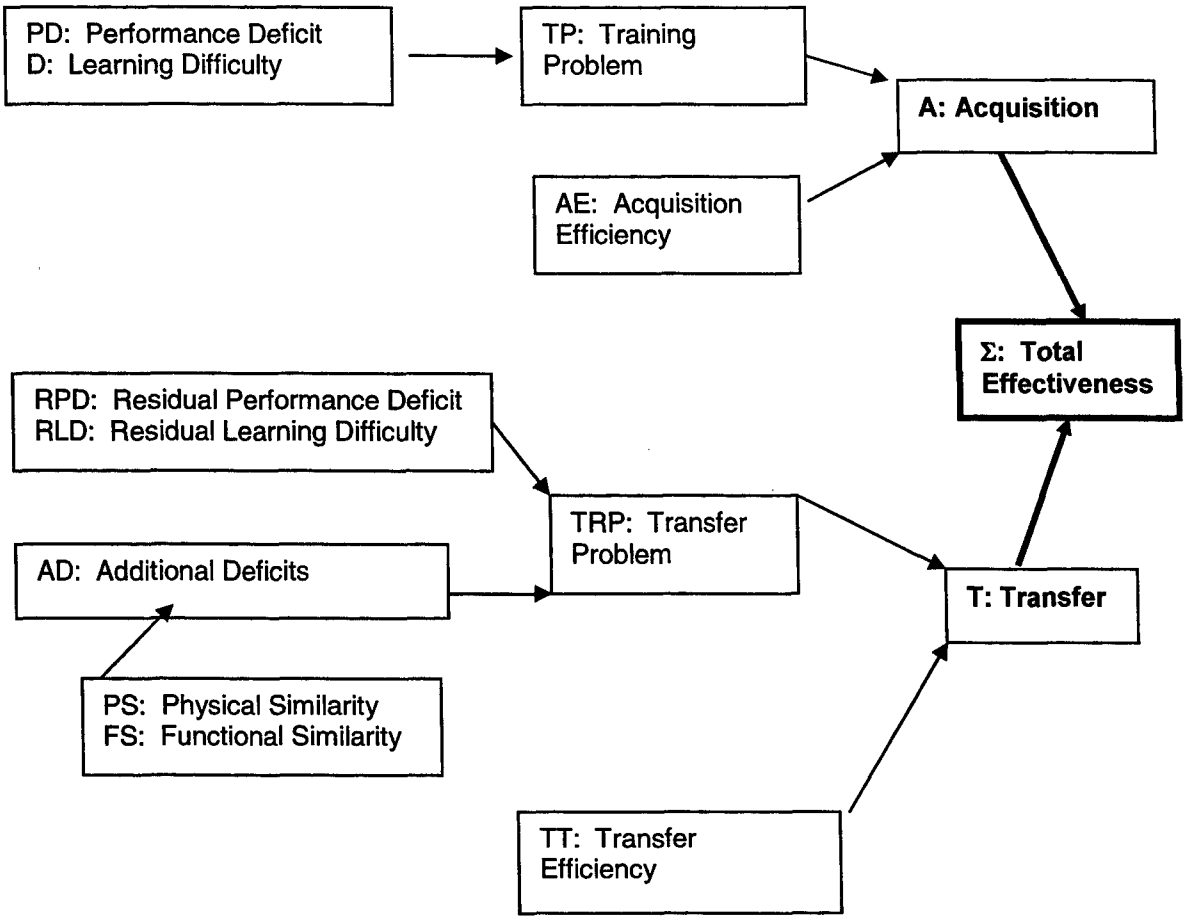


Figure 1. Analysis components of the DEFT and the Training Technology Evaluation Tool.

Method

Development of the Training Technology Evaluation Tool

Scope

The DEFT was the basis for the Training Technology Evaluation Tool. One advantage of the DEFT was its three levels of analysis, which provided flexibility in assessments. It was determined that these three levels could also apply to evaluating training technologies that were in the design phase (Level I), the prototype phase (Level II), and full production and fielded phase (Level III). Consequently, the Training Technology Evaluation Tool as presented here was split into the three distinct tools corresponding to Levels I, II, and III.

The DEFT was originally designed to evaluate training simulators/devices. Since it was based on task and subtask evaluations of pre- and post-training deficits, training methodologies, and educational features, the potential to broaden its application was high. The Training

Technology Evaluation Tool described here was also applied to interactive multimedia instruction (IMI) and to gaming technology (virtual reality/virtual simulations) to determine its ability to render meaningful assessment results. Since most IMI does not allow for realistic, tactile interface with real equipment controls and displays, modifications were made to the Physical and Functional Similarity scales related to equipment controls and displays. However, the Level III analysis requires an evaluation for each of these areas by display or control, and a key element of the design for the Tool was a "one size fits all" approach. After initial testing, it was determined that when performing an evaluation of a training technology that has neither controls nor displays, certain standard numbers could be applied that would not adversely impact the results of the evaluation.

Determining the Tasks and Subtasks to Evaluate

During the review of the DEFT and in the process of determining how to make the Training Technology Evaluation Tool functional for SMEs, it became clear that the SME would have to determine what tasks and subtasks to evaluate. It also became clear that a complex training technology, such as a simulation, might require more than one evaluation based on the total number of tasks and subtasks required for evaluation. For this reason and to avoid confusion, the Training Technology Evaluation Tool limited the number of major tasks to one, and the subtasks to a total of 20. This also allowed greater focus on critical tasks and the possible means by which the training technology could be improved.

Modifications to the DEFT

Automation. The Training Technology Evaluation Tool required automation, as the formulae in DEFT were too complex to calculate manually. The original DEFT was programmed in the common business operating language (COBOL), but the COBOL source code for the DEFT was not available. Therefore, a different software language was selected for the Tool. Microsoft Excel® was selected as the program in which to develop the Tool, due to Excel's flexibility, availability, and ease of use. Excel does not require any additional technical or programming skills, and it is commonly available in the software suite Microsoft Office®, which increases its value to the Army. Excel can run on desktop personal computers, laptop computers, and personal digital assistants (PDAs). In the case of a PDA, the amount of text in the evaluation Tool would require that questions either be modified to fit on the screen, or transferred to a paper worksheet. The entries and subtasks in the electronic spreadsheet would then require cross-referencing to the paper worksheet. Use with a PDA was not tested during the research project.

Separate spreadsheets were developed for each level of analysis. Within each spreadsheet, the questions pertaining to each dimension were contained on different worksheets. Several computational changes were made. Not all changes are cited here, but they are typical of the type of change made to the Performance Deficit scale described below.

During revisions to the Tool, it was discovered that if the Performance Deficit was rated as zero (0) for all subtasks, the entire acquisition score for the technology would result in zero (0). This situation was created by the mathematical rule of zero times any given number, which

will result in the sum of zero. In order to resolve this potential problem, the lowest possible score for each subtask with a fully trained student was changed from zero to one (1). This same problem was encountered with the scales that ranged from 0 to 100. To resolve the problem the 0 was dropped and the scale was reset to 1 to 100. These changes did not affect the formulae, only the allowable range of numbers that the rater could enter into the response fields for the questions.

The final formulae are at Appendix G. With minor exceptions they are the DEFT formulae. Descriptions of the formulae were added to clarify their meaning and linkages.

Reading comprehension level. The questions in DEFT were worded awkwardly and revisions were needed. A behavioral scientist tested the Level I aspect of the draft version of the evaluation Tool, and indicated that the questions had to be read two to three times to fully understand them. This feedback resulted in the revision of some of the questions for all three levels, and then re-testing with two retired non-commissioned officers. Additional minor revisions to the questions were made, and a third review was conducted by a second behavioral scientist. The third review resulted in the deletion of an instruction sheet at the beginning of each level. These instructions were moved to the head of each pertinent set of scales within each level, allowing the raters to review those instructions while performing the evaluation of those particular aspects of the training technology.

In this revision process, the meaning of the questions was not altered, only the reading comprehension level required to understand them and to answer them appropriately. For example, in the first question in Level I that pertains to Transfer Problem Analysis, the question was modified. The original question was, "Question 1. What proportion of the enabling skills and knowledges required in order to reach criterion proficiency on the operational equipment will the trainee still have to learn?" The modified question became, "Question 1. Examine the training objective(s). Considering what you know about the average Soldier's background and training, what percentage of the skills and knowledges will the Soldier still need to learn in order to reach proficiency in the training objective?"

As another example, in Level II the second question pertaining to the Determination of Performance Deficit, read, "Consider each task or subtask that you indicated a trainee won't be able to perform initially on the training device. Rate the difficulty the typical trainee will have in learning to perform each task or subtask." The question was reworded as follows, "Consider the enabling skills and knowledge required to meet the training objective(s) that the typical Soldier does NOT possess. Rate the difficulty of acquiring the remaining skills and knowledges."

In all questions, either "Soldier" or "Soldier to be trained" replaced the word "trainee." Most training technologies will be used in unit training, making the word "trainee" inappropriate.

Application to non-simulator training media. In determining the Residual Learning Difficulty in Level III, the DEFT had specific questions about simulator controls and displays. Since training technologies other than simulators and emulations do not always have displays and controls, a way to complete this portion was required that would not affect the overall score either positively or negatively. Simply entering zero would skew the results and give an

inaccurate assessment of the technology in the area of transfer. After several trials, it was found that the total score for Residual Learning Difficulty was not changed if a Physical Similarity score of 4 was given, a Functional Similarity score of 2 was given, and the total number of displays and controls was assigned a score of 1. The instructions to the rater for the Residual Learning Difficulty controls and displays section indicate the scores to place in the required areas if the training technology under evaluation is not a simulator.

Development of a Scoring Guide

A Scoring Guide was added to the Tool to help the user to interpret the output. Several approaches were examined in this process. One approach simply involved providing the overall minimum and maximum values for a basis of comparison. However, this provided insufficient information on critical dimensions, did not reflect the sophistication of the evaluation Tool, and therefore was assessed as inadequate. Another procedure involved jointly categorizing scores on the basis of the learning difficulty of the tasks, the typical degree of skill that Soldiers had on the tasks prior to using the training technology, and the extent to which the technology incorporated principles of training. This resulted in too many combinations of scores for the user and was also rejected. A third approach focused on summarizing how much could be learned by using the technology, but this technique appeared to ignore other critical dimensions embedded in the Tool. Lastly, it was clear during development of the Scoring Guide that categorizing training technologies into general classes such as good, average, or poor would be an inadequate representation of any technology.

The solution to the Scoring Guide was to present the scores on all the major dimensions, the maximum and minimum scores, whether a low score or high score was desired, and the ratio of each score to the corresponding maximum score to obtain a common comparison index. The general questions addressed by the dimensions under the major components of Acquisition and Transfer are cited below.

Acquisition

- The nature of the training challenge. This is called the Training Problem.
    - How much does the Soldier already know about the task and subtasks? This relates to the Performance Deficit dimension.
    - How difficult is each task or subtask to train? This relates to the Learning Difficulty dimension.
- How well does the training technology use educational principles? This relates to the Acquisition Efficiency dimension.

Transfer

- Transfer of skill from the technology to the actual system. This is called the Transfer Problem.
    - How much more will the Soldier still need to know once he/she can perform all the tasks and subtasks that the training technology presents? This relates to the Residual Performance Deficit dimension.

- How difficult are the remaining tasks and subtasks to train in the real world after using the training technology? This relates to the Residual Learning Difficulty dimension.
- How realistic is the training technology? What are the similarities in physical layout and function? This relates to two Additional Deficits dimensions that assess Functional and Physical Similarities.

- What is the extent of practice a Soldier can perform on the device, and is it realistic and relevant? This relates to the Transfer Efficiency dimension.

In the Scoring Guide, data are presented on each of the above dimensions plus total effectiveness. A total of 14 scores is presented to help the user assess the training technology. Each score is described in detail in the sections that follow. All results were computed automatically in Excel®, eliminating any requirement on part of the user to convert results for comparison purposes.

Description of the Training Technology Evaluation Tool

Levels of Evaluation

The Training Technology Evaluation Tool, like the DEFT, has three distinct levels of evaluation. Each serves its own purpose as indicated below:

- Level I – evaluate newly emerging training technology designs. Very little information about the training technology and the skills and knowledge concerning the subtasks is required. Level I does not evaluate subtasks and is very general in nature.
- Level II – evaluate prototype training technologies, where the final design and development may not yet be completed. Level II does evaluate some aspects of subtasks. More specific information about the training technology and the skills and knowledge concerning the subtasks is necessary to complete Level II, which should normally be available in the testing of a prototype.
- Level III – evaluate fully developed and current training technologies. Very specific information about the training technology and about the skills and knowledge concerning the subtasks is required.

Each of the levels is executed in Microsoft Excel®, and each consists of seven worksheets. Each worksheet covers a specific area of the training technology under evaluation. The worksheets are:

- Training Objectives and Subtasks
- Training Problem (TP)
- Acquisition Efficiency *(AE)*
- Transfer Problem *(TRP)*
- Transfer Efficiency *(TT)*
- Scoring Guide
- Data Sheet

Each workbook and worksheet are protected, only allowing entries where entries should go. It is not possible to erase or modify the Tool without turning off the protection for each worksheet.

Training Objectives and Subtasks

The first worksheet, Training Objectives and Subtasks, limits the subtasks to a maximum of 20. The Tool can be expanded to include more substasks, but the likelihood of this requirement is low. In addition, a longer list of subtasks would probably be quite heterogeneous, and a better approach would be to apply the Tool to distinct groups of subtasks. Evaluating fewer subtasks does not affect the Tool's results for Level I. In Levels II and III the number of subtasks does impact the scores, derived from the Performance Deficit, Acquisition Efficiency, Transfer Problem, and Transfer Efficiency dimensions.

On the first worksheet in the program the rater or SME must enter the overall training objective, each of the subtasks, and the total number of subtasks that constitute the overall training objective. These steps are most critical in Level III, where each subtask is evaluated in each area, and the total number of subtasks to be trained is used in a mathematical formula to derive the Residual Performance Deficit score. Where the subtasks are evaluated individually for a dimension in Levels II and III, they are listed automatically from the Training Objectives and Subtask spreadsheet for Levels II and III. Level I does not consider the number of subtasks or rate individual subtasks.

Relationships Among the Evaluation Dimensions

Figure 1, presented previously, illustrated the relationships among the analysis components (evaluation dimensions) of the Training Technology Evaluation Tool. Each dimension and the questions used at each level of analysis are presented below.

Training Problem (TP): Performance Deficit (PD) and Learning Difficulty (D)

The second worksheet, Training Problem, determines the Soldier's deficit in performance and knowledge prior to training with the technology being evaluated. Answers are subjective, and rely on the expertise of the SME to make the determination. The second part of the spreadsheet asks the SME how difficult the tasks are to train. Again, this is subjective, relying on the rater's expertise to answer appropriately.

As the levels progress, the questions become more specific, and are more directly related to each subtask. For example, Level I contains two questions, one about Performance Deficit, and one about Learning Difficulty. Level II is similar, but the Learning Difficulty question requires an evaluation of each subtask. Level III contains one question concerning Performance Deficit that requires an evaluation of each subtask, and six questions concerning Learning Difficulty, each requiring an evaluation of each subtask.

The Performance Deficit question in Levels I and II is rated on a percentage scale, asking what percentage of the subtasks the Soldier will need to be trained on. In Level III, the

Performance Deficit is based on a scale in which the rater must determine the level of Soldier proficiency for each subtask. The Level III questions are quite specific, requiring a good understanding by the rater of both the subtasks to be trained and the training technology. Questions on the Performance Deficit and Learning Difficulty dimensions for each level are summarized in Table 1.

Table 1
Summary of Questions on Training Problem (TP)

Level I	Level II	Level III
Performance Deficit (PD)		
..Considering the training objective and the Soldier's background, what % of required skills & knowledge will the Soldier need to learn to reach criterion proficiency in the training technology? Scale: 1 (None - can already meet the training objective) to 100 (All – must learn all skills & knowledge)	Same question and scale as Level I	.. Considering Soldier background, rate the current level of proficiency on each subtask. Scale: 10 (no experience); 9 (limited knowledge), 7 (had a complete briefing on subtask / familiarization); 4 (can perform but needs more practice); 1 (can do subtask completely & accurately without supervision).
Learning Difficulty (D)		
Rate the difficulty the typical Soldier will have in acquiring these skills and knowledge. Scale: 1: (Very easy to learn; requires no practice/training) to 100 (Very difficult to learn; requires much training/practice)	Same question and scale as Level I, but SME must rate the difficulty of learning each subtask.	Six questions, answered for each of the subtasks with regard to performance using the training technology. Scale: lower score means less demand on the Soldier 1. Are job/memory aids used in the training technology? Scale: 0 if memory aids; 1 if no memory aids. 2. How many steps required? Scale: 0 if < 10 steps; 1 if > 10 steps 3. Must steps be performed in a definite sequence? Scale: 1 if sequence required; 0 if sequence is not critical. 4. Does subtask have a natural (vs arbitrary) logic to the steps? Scale: 0 if built-in logic; 1 if no built-in logic 5. What are the mental requirements? Scale: 0 if few mental requirements; 3 if mentally challenging. 6. What are motor control demands of the subtask? Scale: 0 if small motor demands; 3 if great motor demands.

Acquisition Efficiency (AE)

The third spreadsheet is on Acquisition Efficiency, which addresses how well the training technology uses sound educational principles. Level I asks only one question, which is very general in nature and requires the rater to judge how well the technology incorporates these principles, on a scale of 1 to 100. The question provides some key indicators to look for in the technology to make this determination. Level II asks four questions, each of which must be rated on a scale of 1 to 100, concerning how well the technology states the performance criteria, provides practice opportunities for the Soldier, provides qualitative feedback to the Soldier, and how well the technology records Soldier performance. Levels I and II Acquisition Efficiency questions are not directly tied to each subtask.

Level III is tied directly to each subtask and asks 11 questions as shown in Table 2. The 11 questions deal with basic principles of psychology that have been shown to facilitate acquisition of skill and provide quality training.

Table 2
Summary of Questions on Acquisition Efficiency (AE)

Level I	Level II	Level III
... Considering four basic instructional features (clear training objective, immediate performance feedback, practice on hard to learn skills, and record of performance) and the performance deficits identified, rate how well use of the training technology will overcome these deficits. Scale: 1 (Poor training, few sound training principles in the training technology) to 100 (Excellent training, use of sound training principles in the training technology)	1. For what % of subtasks to be learned does the training technology make the criterion performance requirements explicit to the Soldier? Scale: 1 (None) to 100 (All)	11 questions. Rate how well the training technology will overcome the performance deficits based on its instructional features. Each subtask is rated. Scale: 1 (Not include or address) to 100 (Completely include or address)
	2. For what % of subtasks that must be learned does the training technology provide practice? Scale: 1 (None) to 100 (All)	1. Extent to which training objectives are made clear and clarify Soldier's proficiency with regard to the objective. 2. Extent to which Soldier progresses from easy to hard components/aspects of subtask.
	3. For what % of the subtasks that must be learned does the training technology provide qualitative feedback? Scale: 1 (None) to 100 (All)	3. Provide knowledge of results & positive reinforcement. 4. Provide repetition & practice
	4. What % of the subtasks that must be learned does the training technology provide a record of Soldier performance? Scale: 1 (No records for any subtasks) to 100 (Records for all subtasks)	5. Prompts early in training and gradually fade prompts as training progresses. 6. Chunk material appropriately to complexity of task. 7. Use of memory aids when appropriate or needed 8. Begins with wide tolerance for correct performance and narrows as

Level I	Level II	Level III
		training progresses. 9. Examples provided of all conditions under which the subtask is performed. 10. Extent to which the subtask is performed under different circumstances. 11. Extent to which technology adapts training features as a function of the Soldier's performance.

Transfer Problem (TRP): Residual Performance Deficit (RPD, Residual Learning Difficulty (RLD), and Additional Deficits (AD) of Physical Similarity (PS) and Functional Similarity (FS)

The fourth spreadsheet is on the Transfer Problem and consists of questions concerning Residual Performance Deficit and Residual Learning Difficulty, plus additional deficits related to Physical and Functional Similarity. The term "residual" is used as it reflects the tasks not addressed or fully trained by the training technology but which must still be acquired, i.e., the remaining or residual requirements. Residual Performance Deficit reflects the size of the Performance Deficit that will exist with respect to operating the equipment or functioning in the real world after the Soldiers have reached the performance criterion in the training technology. Residual Learning Difficulty is the difficulty of learning the remaining subtasks or their components in the operational or real world situation. Lastly, Additional Deficits (AD) addresses the physical and functional differences between the training technology and the operational equipment.

Level I contains only a single question for each of the four dimensions or components that constitute the Transfer Problem. First, SMEs are asked what percentage of the tasks, on a scale of 1 to 100, remain to be trained to performance criterion in the real world after the Soldier has reached the performance criterion on the training technology. The second question in Level I pertains to how difficult it will be to train the Soldier on those remaining tasks, on a scale of 1 to 100, in the real world situation. The third question is directed at the physical similarities that exist between the training technology and the real world situation, on a scale of 1 to 100. The fourth and final question in Level I is directed at the functional similarities between the training technology and the real world situation, on a scale of 1 to 100.

The questions in Level II on the Transfer Problem are more complex. The question concerning Residual Performance Deficit asks the evaluator to rate Soldiers being trained on their ability to perform the residual subtasks depending on whether or not these subtasks were presented in the training technology, and whether or not the Soldiers can accomplish the subtasks. There are four answers, scaled 1 through 4, and the evaluator must answer the question for each subtask listed. Residual Learning Difficulty is gauged on a scale of 1 to 100 for each subtask listed. Physical and Functional Similarities between the training technology and the operational situation are also evaluated on a scale of 1 to 100 for each subtask.

Level III's Transfer Problem is evaluated with nine questions. The first question pertains to the Residual Performance Deficit and the ability of the Soldier to perform each subtask listed. The next six questions are on Residual Learning Difficulty. Individual questions concern job aids or memory aids, how many steps are required to perform the subtasks, whether or not each of the subtasks must follow a specified sequence in their performance, and if there is a natural logic to help the Soldiers determine if they are performing the subtasks correctly. These questions are rated either 1 or 0 – they do or they do not have these qualities. The last two questions pertaining to Residual Learning Difficulty ask how mentally challenging the subtasks are, and what motor skills are involved in mastering the subtasks. These two questions are rated as either 0 or 3. A score of 0 indicates the subtask is not mentally challenging or it does not require refined motor skills. A score of 3 means the opposite.

At Level III, the last two questions under Transfer Problem concern the Physical and Functional Similarities between the training technology and the operational situation, but are based on every control and display, and on how closely each resembles and works like the real world situation. The questions are similar to those used for Level II, but each control and display must be assessed, as opposed to providing a rating for each subtask. The last requirements for input ask for the number of displays, the number of controls in the operational situation as well as a short description of each. When the evaluator completes this spreadsheet for a non-simulator, instructions are given to complete the questions with pre-determined responses, as indicated in Table 3. Table 3 presents the questions that generate the Transfer Problem score for each level of analysis.

Table 3
Summary of Questions on Transfer Problem (TRP)

Level I	**Level II**	**Level III**
Residual Performance Deficit (RPD)		
...What % of skills & knowledges must the Soldier learn on the real world equipment/situation because they are not taught or practiced with the training technology? Scale: 1 (None, Soldier already possesses the skills and knowledges not covered by the training technology) to 100 (All; must learn all the skills and knowledges not covered).	For each subtask, rate the extent to which it is covered by the training technology and how difficult it is to learn on the operational equipment. Scale: 1 (Subtask in training objective; easy to learn on actual equipment or in real world) 2: (Subtask not in training objective: easy to learn) 3: (Subtask in training objective: hard to learn on actual equipment or in real world) 4: (Subtask not in training objective; hard to learn)	Assume the Soldier can perform all subtasks comprising the training objective in the training technology. Rate the extent to which the Soldier can perform each subtask on the operational equipment or real world. Scale: 10 (No background, cannot perform) 9 (Limited knowledge) 7 (Can perform if assisted with each step) 4 (Can perform with supervision; needs more training) 1 (Can perform completely and accurately without supervision)

Level I	Level II	Level III
Residual Learning Difficulty (RLD)		
.. Rate the difficulty of acquiring the skills and knowledges in the real world that are not covered by the training technology. Scale: 1 (Very easy to learn, no training or practice needed) to 100. (Very difficult to learn, much training or practice needed)	Rate the difficulty the typical Soldier would have in learning to perform each subtask in the real world after being trained with the training technology. Scale: 1 (very easy to learn) to 100 (very difficult to learn)	For each subtask, rate the performance requirement on the operational equipment or real world. 1. Are there job or memory aids to assist with performance? Scale: 0 if memory aids; 1 if no memory aids 2. How many steps are required? Scale: 0 if < 10 steps, 1 if > 10 steps 3. Must steps be done in a specific sequence? Scale: 1 if sequence required; 0 if sequence is not critical 4. Does the subtask have a built-in logic? Scale: 0 if built-in logic; 1 if no built-in logic 5. What are the mental requirements of the subtask? Scale: 0 if few mental requirements; 3 if mentally challenging 6. What are the motor control demands of the subtask? Scale: 0 if small motor demands; 3 if great motor demands.
Additional Deficits (AD)		
Physical Similarities (PS)		
Rating of physical similarity between training technology and operational equipment. Scale: 1 (Totally dissimilar) to 100 (Identical)	Same as Level I, except each subtask must be rated.	Rating of physical similarity for each control/display. Scale: 1 (Totally dissimilar) to 100 (Identical) For training technology that is not a simulator, enter a 4.
Functional Similarities (FS)		
Rating of functional similarity (Information flow from display to user and from user to controls, and type of information processing) between training technology and operational equipment. Scale: 1 (Totally dissimilar) to 100 (Identical)	Same as Level I, except each subtask must be rated.	Rating of functional similarity for each control/display. Scale: 1 (Totally dissimilar) to 100 (Identical) For training technology that is not a simulator, enter a 2.

Transfer Efficiency (TT)

The fifth spreadsheet, Transfer Efficiency, evaluates the training technology in the area of education and training principles that will assist the Soldier in transferring the skills mastered in the technology to the operational situation. In Level I, there is just one question on the degree to which transfer is facilitated by the technology using a scale of 1 to 100. Level II has three questions, each on a scale of 1 to 100, which concern realism, the amount of practice available in

the training technology, and how well the exercises relate to transferring skills to the real world situation. Level III has only one question, rated on a scale of 1 to 100, which pertains to amount of practice provided to enable Soldiers to perform each subtask in the real world. Table 4 summarizes the questions on Transfer Efficiency.

Table 4
Summary of Questions on Transfer Efficiency (TT)

Level I	Level II	Level III
...Considering the instructional and training features in the technology, rate how well the technology will promote transfer to the real world situation. Scale: 1 (Poor transfer, few instructional features to promote transfer) to 100 (Excellent transfer, maximum use of instructional features to promote transfer)	1. What % of subtasks to be learned in the technology are realistic, relevant, or similar to subtasks performed in real world? Scale: 1 (No subtasks) to 100 (All subtasks)	Consider the instructional features of the training technology that increase transfer. For each subtask, to what extent will the technology permit Soldiers to practice until they can demonstrate a job entry skill level? Scale: 1 (Not at all, no amount of practice will not enable them to perform at job entry level) to 100 (Completely; will be at job entry level).
	2. For what % of subtasks to be learned in the technology are the conditions of practice presented late in training similar to those in the real world? Scale: 1 (None) to 100 (All)	
	3. For what % of subtasks that must be learned is an extensive amount of practice given in the training technology? Scale: 1 (None) to 100 (All)	

Scoring Guide

As cited previously, a Scoring Guide was developed to help the rater interpret the scores on the primary dimensions in the Training Technology Evaluation Tool. The Scoring Guide is the sixth and last spreadsheet in each of the levels' workbooks. There is no requirement for input to this spreadsheet, as all input is derived from the other spreadsheets. It is completely automated and protected. The Scoring Guide is described in the Results section.

Data Sheet

The last worksheet, the Data Sheet, is also completely automated and protected, and is where the necessary calculations are performed. It serves only as a data work and storage area.

Documentation

The final versions of the Training Technology Evaluation Tool are found at Appendix D for Level I, Appendix E for Level II, and Appendix F for Level III. Formulae and mathematical equations used to derive scores are discussed in detail in Appendix G, Formulae.

Assessment of the Training Technology Evaluation Tool

Examination of Rater Preferences

The questions and the formulae together suggested that an evaluation performed by a SME who might possess a preference for or bias against a particular training technology could be neutralized when specific questions were used. Level III of the Training Technology Evaluation Tool provided the context for assessing this concern, as the questions were the most specific of the three levels, and almost all questions were directed against each subtask.

Different reactions by former military personnel to an interactive training game designed to train squad leaders presented an opportunity to test the hypothesis that specific, rather than general, evaluation questions would reduce bias. Two retired senior non-commissioned officers who had used the game had diametrically opposed opinions of the game as an effective training technology. They were requested to conduct an evaluation of the game utilizing the Training Technology Evaluation Tool with the Level III analysis. The results of the evaluation were compared to their anecdotal evaluations to determine how well the questions and structure of the Tool reduced or eliminated their biases concerning the game as a training technology. The results of this comparison are discussed under the Games for Training section.

Application to Different Training Technologies

One of the requirements of this research project was to evaluate four training technologies. Of those four, some were required to be established or legacy training technologies, and some were required to be emerging training technologies. During the evaluations of the training technologies, every effort was made to secure SMEs in the fields covered by the technologies. Prior to SMEs using the evaluation Tool, they were questioned about their level of experience in the related field. They were also questioned concerning any positive or negative biases they may have toward a particular training technology. Upon completion of the evaluation, they were questioned about some of their answers in order to clarify the rationale for their responses. When the SMEs used the Tool, they were not interrupted nor allowed to ask questions, in an effort to ensure that the instructions within the Tool were adequate and understandable.

Gunnery simulator – Levels II and III. The Unit Conduct of Fire Trainer (UCOFT), an established or legacy technology, is a training simulator requiring an instructor/observer and is designed to train the Bradley Fighting Vehicle Commander and gunner how to accurately engage targets in a variety of battlefield scenarios in both temperate and desert climates. It is a mock up of the turret and all fire controls that would be encountered in an actual Bradley Fighting Vehicle. The UCOFT was evaluated using Levels II and III of the Training Technology Evaluation Tool. A retired senior non-commissioned officer with extensive experience using the UCOFT to train Soldiers made both evaluations.

IMI – Levels I, II and III. An IMI program on map reading, used to train prerequisite skills for a Soldier-based digital system, was selected for evaluation. This is also an established technology. All levels of analysis were used to assess the IMI. Since Level III is oriented to the

evaluation of controls and displays normally found in a simulator, it was necessary to provide pre-determined answers for control and display questions, as discussed in the prior section on modifications to the DEFT. The rater was a retired officer with extensive experience in training Soldiers, teaching map-reading skills, and designing and developing interactive multimedia instruction.

Individual weapons simulator – Levels I and II. The fourth evaluation was conducted on an individual weapons training simulator, classified as an emerging technology, designed to help Soldiers learn basic rifle marksmanship (BRM) and portions of advanced rifle marksmanship (ARM). The training technology also provides selected combat marksmanship training in a synthetic environment. This element of the simulator provides combat scenarios for rifle squads, enabling them to control fires and engage targets as a team. This element of the technology was not evaluated. The evaluation of subtasks concerned only BRM and ARM as outlined in FM 3-22.9, Appendix F, Actions, Conditions, and Standards (Department of the Army, 2003). The key problem involved with this evaluation was the fact that the actions, conditions, and standards for ARM also include firing the M16 or M4 rifle using optics and sights (close combat optic, the thermal weapon sight, and the infrared aiming light with night vision goggles) that are not currently incorporated in the individual weapons simulator. The simulator only trains Soldiers how to use open iron sights. These subtasks, although inherent to ARM, were not were included as part of the evaluation as they constituted a large percentage of the subtasks (40%, 8 of 20). Three research psychologists who had experience with the simulator made the evaluations.

Games for training – Level III. The Army is looking more and more to private industry and to three-dimensional gaming technologies (virtual simulations) to support training. Gaming technology was considered an emerging training technology. The game that was assessed had been used in a prior experiment on training military operations in urban terrain. Two infantry SMEs who had performed experiments with the game were requested to apply the Tool to assess its training value. As mentioned previously, since there was anecdotal information that the two SMEs had divergent opinions of the value of the game for training purposes, this particular assessment also provided an opportunity to determine if the Tool could reduce rater preference or bias.

Results

As stated in the Method section, the Training Technology Evaluation Tool was assessed from several perspectives. It was used to assess the training effectiveness of four training technologies. A Scoring Guide was developed to assist SMEs in determining how well the evaluated training technology performed as compared to a standard. The Scoring Guide is presented first, since it is used to show the results of the evaluations of the four training technologies.

Scoring Guide

Scoring Guide format. The format for the Scoring Guide is shown in Table 5. All computations for the guide are fully automated. Within the Excel program for the Training Technology Evaluation Tool, the Scoring Guide results can be found at the Scoring Guide worksheet. When the SME completes the evaluation, the Scoring Guide is also complete.

Table 5
Format for the Scoring Guide

Score Comparisons, Level I, II or III					
Name of Training Technology					
			# of Subtasks		
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	Derived			Lower is better	
-Acquisition	Derived			Lower is better	
-Transfer	Derived			Lower is better	
Training Problem	Derived			Lower is better	
-Performance Deficit	Rating			Score is neutral	
-Learning Difficulty	Rating			Lower is better	
Acquisition Efficiency	Rating			Higher is better	
Transfer Problem	Derived			Lower is better	
-Residual Performance Deficit	Rating			Lower is better	
-Residual Learning Difficulty	Rating			Lower is better	
-Additional Deficits	Derived			Lower is better	
--Physical Similarities	Rating			Higher is better	
--Functional Similarities	Rating			Higher is better	
Transfer Efficiency	Rating			Higher is better	

The top section of the guide documents the level of comparison, the training technology evaluated, and the number of subtasks that it trains. The number of subtasks is used to calculate the maximum and minimum scores for Levels II and III.

The first column, named *Dimension/Component,* lists the major dimensions by which each training technology is evaluated. These correspond to the dimensions presented in Figure 1. They are in three groups. The first group presents the three summary dimensions: Total Effectiveness and its two components of Acquisition and Transfer. The second group contains the scores that compose the Acquisition dimension. The last group contains the scores that compose the Transfer dimension.

To the right of the *Dimension/Component* column is *Actual Scores Received During Evaluation.* Some dimensions are labeled "Rating" and others are labeled "Derived." The term

"Rating" means that only the SME's ratings for that particular dimension are used to determine the score for the dimension. On the other hand, the term "Derived" means the score on that dimension is based on scores from at least two other dimensions. Appropriate interpretation of the score for a derived dimension requires knowledge of the formula used for its calculation.

The next column is *Maximum Possible Score*, which represents the highest score possible for each dimension. To the right of the *Maximum Possible Score* is the *Minimum Possible Score*, showing the exact opposite of the previous column. The maximum and minimum scores for each dimension or component represent the possible range of scores for the technology.

The column to the right of *Minimum Possible Score* is *Score Interpretation.* This column indicates if a high or a low score is desirable in the evaluation for each dimension or component. Note that many dimensions are inversely scaled, where low values represent the desired direction (e.g., high effectiveness or high transfer).

The final column is *Percent of Maximum Score*, which is the ratio of the score the rater gave the technology for that dimension or component to the maximum possible score, expressed as a percentage. Again, because it is based on the maximum possible score, the Score Interpretation can be applied directly to the percentage as well. In other words, if a high score is desirable and the component scored a high percentage, the better the technology. If a lower score is desirable, the lower percentage is also desirable. For most dimensions in the Training Technology Evaluation Tool, a low percentage score is desirable. Performance Deficit scores are described as "neutral," as the typical Soldier could know much about the task prior to training or very little. The central issue is whether the training technology adequately addresses the deficits that exist initially, not the size of the initial performance deficit itself.

Table 6 presents a sample Scoring Guide output. As indicated in the top section of the guide output, this was a Level III evaluation of an IMI training program with ten subtasks.

As shown in Table 6, the SME assessed the Performance Deficit as high. As stated previously, a score of 99% of the maximum possible score on Performance Deficit does not mean that the training technology is flawed. It means that the rater believes that Soldiers to be trained with the technology initially know very little if anything about the task and subtasks to be learned. As indicated by the Learning Difficulty percentage score of 24% (lower score is better), the rater felt the tasks to be learned were not difficult to learn. Consequently, the Training Problem received a desirable (low) percentage score (24%), even though the initial performance level was perceived as being low. Also, as indicated by the Acquisition Efficiency percentage score of 85% (higher score is better), the training technology was viewed by the rater as making good use of educational principles.

With regard to transfer, both the Residual Performance Deficit and the Residual Learning Difficulty received relatively low percentage scores. This indicates that the training technology adequately covered most of the subtasks; those aspects of the task that were not addressed were perceived as not difficult to learn. Because IMI technology is not a simulator, by default it received low similarity scores. If the training technology evaluated had been a simulator, then high scores would be desired on the Physical and Functional Similarities dimensions, assuming

these dimensions are essential to high transfer of skill. The SME assessed the technology's capability to facilitate transfer of skill to the real world as good, but not the best. The percentage score for Transfer Efficiency was 71%, but a higher score is desirable.

Table 6
Example of Level III Scoring Guide Results (IMI)

Score Comparisons, Level III					
Name of Training Technology **Sample Level III**					
		10	# of Subtasks		
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	395.74	20990.00	0.00	Lower is better	2%
-Acquisition	280.64	10000.00	0.00	Lower is better	3%
-Transfer	115.10	10990.00	0.00	Lower is better	1%
Training Problem	237.60	1000.00	0.00	Lower is better	24%
-Performance Deficit	99.00	100.00	10.00	Score is neutral	99%
-Learning Difficulty	24.00	100.00	0.00	Lower is better	24%
Acquisition Efficiency	0.8467	1.00	0.10	Higher is better	85%
Transfer Problem	82.20	1099.00	0.00	Lower is better	7%
-Residual Performance Deficit	28.00	100.00	10.00	Lower is better	28%
-Residual Learning Difficulty	29.00	100.00	0.00	Lower is better	29%
-Additional Deficits	1.00	99.00	0.00	Lower is better	1%
--Physical Similarities	2.00	100.00	1.00	Higher is better	2%
--Functional Similarities	1.00	100.00	1.00	Higher is better	1%
Transfer Efficiency	0.7141	1.00	0.10	Higher is better	71%

Lastly, the summary percentage scores for Total Effectiveness, Acquisition, and Transfer were all low, which is desirable. However, for Level III these three summary percentage scores can easily be very low simply because of the way in which they are calculated. The denominator for each percentage is directly related to the number of tasks assessed. Consequently, the maximum value which is the basis for the summary percentage can be very large. This is illustrated in Table 6. For a Level III analysis, the most important and useful information about the technology is derived from the subscores that generate the overall Acquisition and Transfer numbers rather than the three summary percentage scores.

For all levels of analysis, it is important to point out that the Total Effectiveness actual score is the sum of the Acquisition and Transfer actual scores. One might think that the Total Effectiveness percentage score would be the arithmetic mean of the Acquisition and Transfer percentage scores. However, this is not necessarily the case, as typically the denominators for these two component percentages are not identical. The denominator for the Total Effectiveness percentage score (which is the maximum possible score) is weighted more heavily by whichever component (Acquisition or Transfer) has the larger denominator (i.e., maximum possible score), which in turn affects the mean. However, if the two percentage scores for Acquisition and Transfer are identical, then the Total Effectiveness percentage score will be identical to the two component percentage scores.

Summary percentage scores. Special comments are made regarding the three summary scores: Total Effectiveness, Acquisition, and Transfer. These are derived dimensions (refer to Table 5). A sensitivity analysis of Acquisition scores for the Level I analysis indicated that the range of possible Acquisition scores for typical training technologies is restricted because of the formula used to calculate Acquisition. Acquisition is the Training Problem score divided by the Acquisition Efficiency score. Training Problem scores can range from 0.01 to 100; Acquisition Efficiency scores can range from 0.1 to 1.0. Figure 2 shows Acquisition scores as a function of three scores for Training Problem (100, 50 and 10), and 10 scores for Acquisition Efficiency (from 0.1 to 1.0).

As indicated in Figure 2, Acquisition Efficiency scores would have to be low (ranging from approximately .1 to .6) in order to show substantial differences in the Acquisition score. In other words, in order for a training technology to reflect this range of Acquisition Efficiency scores, it would need to possess few, if any, good principles of training, which is rarely the case. Thus, relatively small differences in Acquisition scores between training technologies in the same task domain may, in fact, reflect practical, important differences.

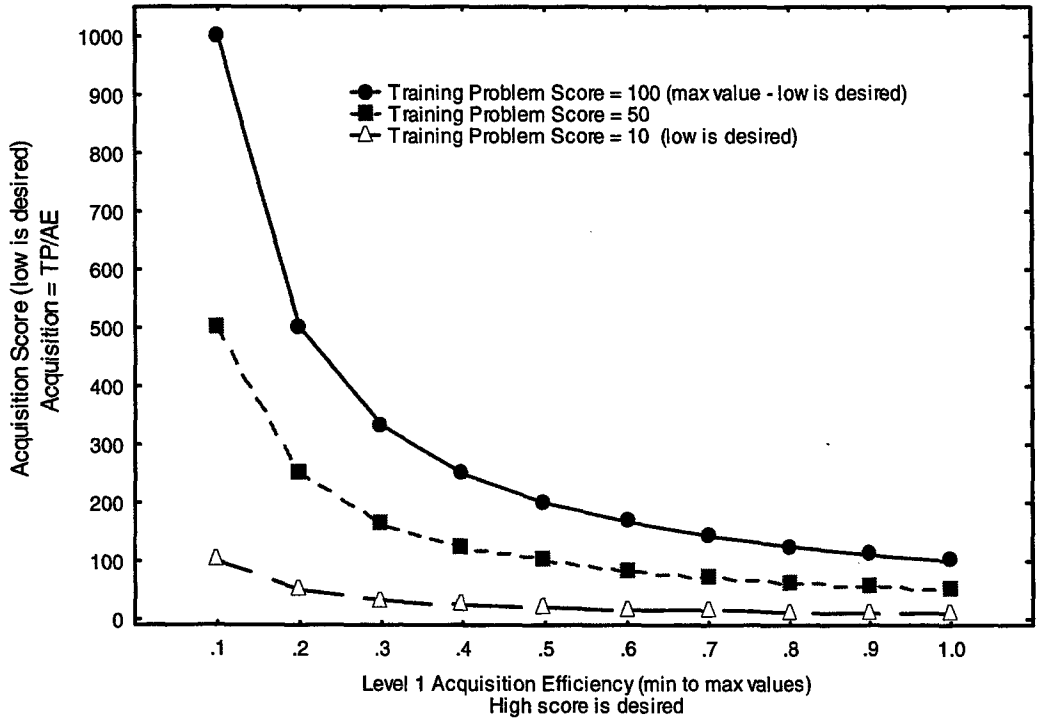


Figure 2. Acquisition score as a function of Training Problem and Acquisition Efficiency scores.

The same relationship applies to how the Transfer score is impacted by the Transfer Efficiency score. The Acquisition and Transfer scores for the Level II analysis are affected similarly. Thus it is very important that the user of the Training Technology Evaluation Tool examine both the actual scores and % of maximum scores.

As stated previously regarding the Level III analyses, the summary component percentages are directly related to the number of subtasks evaluated. The denominator for these percentages (the maximum numbers) can be very large. As such, the technology would have to be rated extremely low for the percentage scores for Total Effectiveness, Acquisition, and Transfer to be above 10%. In the evaluation results presented in this report, no summary percentage scores were above 5%. With the Level III analysis, the other component scores yielded the best insights into the SME's perception of the training technology.

Sensitivity analyses. A sensitivity analysis was conducted with Level I. This analysis compared the results obtained from varying the percentage of tasks trained with the technology, as well as scores assigned to the two efficiency scores (Acquisition and Transfer). This analysis was based on characteristics of the individual weapons simulator which is referenced later in the report. An SME identified 17 individual marksmanship subtasks which a small-arms weapons simulator should/could train. Of these 17, the SME indicated that seven were trained, at least in part, by the current individual weapons simulator (see Table 7). In addition, the SME rated the weapons simulator as incorporating sound training principles for Acquisition Efficiency and Transfer Efficiency (ratings of 70 and 75 respectively on the 100 point scale). Thus, the baseline for the sensitivity analysis involved the seven marksmanship tasks that were trained by the simulator, and assumed that the simulator included sound training principles.

Table 7
Marksmanship Subtasks Used in the Level I Sensitivity Analysis

Subtasks Trained in the Simulator	Subtasks not Trained in the Simulator
Perform function checks	Load/unload M16/M4 magazine
Correct Malfunctions	Detect/engage moving targets
Load/unload M16/M4	Boresight Close Combat Optic (CCO)
Grouping (iron sights)	Zero the CCO
Zeroing (iron sights)	Boresight the aiming light
Detect/engage single stationary timed targets	Boresight the thermal sight
Detect/engage multiple stationary timed targets	Zero the thermal sight
	Detect/engage targets with CCO
	Detect/engage targets with aiming light and night vision goggles
	Detect/engage targets with the thermal sight

Three other variations were examined in the analysis. All four variations in the sensitivity analysis are presented below. The percentage of subtasks trained on the simulator was either 100% (7 tasks) or 40% (17 tasks). The Acquisition Efficiency and Transfer Efficiency Ratings were either good (70 & 75, respectively) or low (20 each). For example, Variations A (baseline) and B used the ratings that the SME thought should be applied to the individual weapons simulator, given what was known about its training features. The only difference between Variations A and B was the percentage of tasks trained in the simulator. Variations C and D also varied the percentage of subtasks trained, but the training feature ratings on Acquisition Efficiency and Transfer Efficiency were low.

Variation	# Subtasks Considered in the Analysis	% Subtasks Trained on Simulator	Acquisition Efficiency and Transfer Efficiency Ratings
• A (baseline)	7	100%	70/75 (good)
• B	17	40%	70/75 (good)
• C	7	100%	20/20 (low)
• D	17	40%	20/20 (low)

Table 8 shows the results of the sensitivity analysis. In general, the conditions that were varied impacted the results.

Table 8
Individual Weapons Simulator Level I Sensitivity Analyses

Dimension/Component	Variation A Baseline All Subasks Trained Current Training Features (AE & TE) **Score / % maximum**	Variation B Some Subtasks Trained Current Training Features (AE & TE) **Score / % maximum**	Variation C All Subtasks Trained Poor Training Features (AE & TE) **Score / % maximum**	Variation D Some Subtasks Trained Poor Training Features (AE & TE) **Score / % maximum**	**Score Interpretation**
Total Effectiveness	101.18 / 3%	195.65 / 7%	193.98 / 6%	372.86 / 12%	Lower is better
-Acquisition	71.30 / 7%	104.21 / 10%	138.08 / 14%	201.82 / 20%	Lower is better
-Transfer	29.88 / 1%	91.43 / 5%	55.90 / 3%	171.06 / 9%	Lower is better
Training Problem	61.75 / 62%	90.25 / 90%	61.75 / 62%	90.25 / 90%	Lower is better
-Performance Deficit	95.00 / 95%	95.00 / 95%	95.00 / 95%	95.00 / 95%	Score is neutral
-Learning Difficulty	65.00 / 65%	95.00 / 95%	65.00 / 65%	95.00 / 95%	Lower is better
Acquisition Efficiency	.8860 / 87% (75 rating)	.8860 / 87% (75 rating)	.4472 / 45% (20 rating)	.4472 / 45% (20 rating)	Higher is better
Transfer Problem	25.00 / 13%	76.50 / 38%	25.00 / 13%	76.50 / 38%	Lower is better
-Residual Performance Deficit	25.00 / 25%	70.00 / 70%	25.00 / 25%	70.00 / 70%	Lower is better
-Residual Learning Difficulty	60.00 / 60%	95.00 / 95%	60.00 / 60%	95.00 / 95%	Lower is better
-Additional Deficits	10.00 / 10%	10.00 / 10%	10.00 / 10%	10.00 / 10%	Lower is better
--Physical Similarities	85.00 / 85%	40.00 / 40%	85.00 / 85%	40.00 / 40%	Higher is better
--Functional Similarities	75.00 / 75%	30.00 / 30%	75.00 / 75%	30.00 / 30%	Higher is better
Transfer Efficiency	.8367 / 84% (70 rating)	.8367 / 84% (70 rating)	.4472 / 45% (20 rating)	.4472 / 45% (20 rating)	Higher is better

The summary scores of Total Effectiveness, Acquisition and Transfer were better for Variation A (baseline), which covered only the subtasks trained and incorporated good ratings on principles of training. The differences in ratings for Variations A and B, where the percentage of tasks trained differed, can be traced to the scores for Learning Difficulty, Residual Performance Deficit, and Residual Difficulty. For Learning Difficulty, the SME assigned poorer ratings in Variation B, as the difficulty of learning all the subtasks was perceived to be rather high. With regard to Residual Performance Deficit and Residual Difficulty, the percentage of skills left to be trained on the actual equipment after achieving mastery on the simulator was higher in Variation B than in Variation A (baseline). These skills were also judged to be more difficult than those that remaining be trained on the actual equipment under Variation A, consequently, there were differences in the SME ratings.

The other two variations, C and D, examined what happened to the overall Training Technology Evaluation Tool ratings when the individual weapons simulator was considered to incorporate poor principles of training. In both instances, a rating of 20 was assigned to Acquisition Efficiency and Transfer Efficiency as opposed to ratings of 75 (AE) and 70 (TE) for Variations A and B. In comparing Variation A to C and Variation B to D, the same result occurred; poor training features received poorer overall ratings, when the rater's assessments on other directly rated dimensions were held constant.

The results for this Level I sensitivity analysis indicated that the baseline (Variation A), where the training technology trained all tasks and incorporated good training principles, had the best scores on Total Effectiveness, Acquisition, and Transfer. On the other hand, the variation where only some of the tasks were trained and the technology had poor training principles yielded the poorest scores (Variation D). As would be expected, the variations with the best training principles (A and B) were rated the best on Acquisition. Variations that trained all the tasks were rated the best on Transfer (Variations A and C), as there were more subtasks remaining to be trained using the actual equipment with the other variations (B and D). Thus the Training Technology Evaluation Tool scores reflected the dimensions that were varied in the sensitivity analysis.

Gunnery Simulator – Levels II and III Evaluation Results

One of the first evaluations conducted with the Training Technology Evaluation Tool was the Unit Conduct of Fire Trainer (UCOFT). The UCOFT has a good reputation as a training system, and presented an opportunity to test the validity of the Tool. Rater 1 evaluated the UCOFT, using both Levels II and III. The overall task was "engage targets with the M2 Bradley Fighting Vehicle 25mm M242 Bushmaster chain gun." The 16 subtasks that support this task and were used to rate the UCOFT are shown in Table 9.

During this evaluation and the evaluation of the marksmanship training device an issue arose that could impact the scoring. The question was whether the training technology should be evaluated on tasks or subtasks that it does not or cannot train. Specifically, the UCOFT does not train the gunner how to reload ammunition for the M242 chain gun. Yet this task is a critical subtask to engaging targets with the M242. The same issue arose with the marksmanship trainer, which does not and currently cannot train Soldiers to engage targets with certain optics and

sights. As indicated in the sensitivity analysis, if the technology is evaluated as if it should train the subtasks, the overall score will not reflect favorably on the technology. If the technology is evaluated as if it should not train all the subtasks, then the score will be more favorable. For evaluation of the UCOFT, subtask 6, reload the prime weapon, was included in the evaluation as if the technology should or could train the subtask. Given that the subtask was only one of 16 subtasks included in the analysis, it was not expected to strongly affect the results.

Table 9
Gunnery Simulator Subtasks

Subtask 1	Use fire commands
Subtask 2	Gunner engages stationary targets from the stationary position using the prime weapon
Subtask 3	Gunner engages stationary targets from the move using the prime weapon
Subtask 4	Gunner engages moving targets from the move using the prime weapon
Subtask 5	Gunner engages moving targets from the stationary position using the prime weapon
Subtask 6	Gunner reloads the prime weapon
Subtask 7	Gunner reloads the coax machinegun
Subtask 8	Commander engages stationary targets from the stationary position using the prime weapon
Subtask 9	Commander engages stationary targets from the move using the prime weapon
Subtask 10	Commander engages moving targets from the move using the prime weapon
Subtask 11	Commander engages moving targets from the stationary position using the prime weapon
Subtask 12	Gunner selects the appropriate ammunition for the target description
Subtask 13	Commander selects the appropriate ammunition for the target description
Subtask 14	Gunner engages targets using the day sight
Subtask 15	Gunner engages targets using the thermal sight
Subtask 16	Gunner correctly identifies the target using the thermal sight

The results for Rater 1's Level III evaluation are shown at Table 10. The Level III percentage and actual scores were generally low where they should have been low, and high where they should have been high. However, as indicated previously, the Total Effectiveness, Acquisition, and Transfer percentage scores can be extremely low when the maximum possible score is very large. This was the case for the gunnery simulator assessment as there were 16 subtasks. Thus, the other component scores in Table 10 provided the best basis for interpreting the rater's perception of the gunnery simulator training technology.

With respect to the Acquisition dimension, the Performance Deficit percentage score (27%) indicated the rater thought the Soldiers to be trained on the gunnery simulator have some background in the subtasks, but are not fully trained. A lower score would have indicated that the Soldiers were perceived to be more fully trained on the subtasks, while a higher score would have indicated that they had little prior experience and training. The percentage score for Learning Difficulty (59%) indicated that the subtasks are moderately difficult to learn using the training technology. The relatively high percentage score for Acquisition Efficiency (77%) indicated the rater felt the UCOFT incorporated sound training principles.

Table 10
Gunnery Simulator Results, Level III Analysis

Score Comparisons, Level III					
Name of Training Technology **Gunnery Simulator**					
	16	# of Subtasks			
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	480.14	32990.00	0.00	Lower is better	1%
-Acquisition	328.10	16000.00	0.00	Lower is better	2%
-Transfer	152.04	16990.00	0.00	Lower is better	1%
Training Problem	252.63	1600.00	0.00	Lower is better	16%
-Performance Deficit	43.00	160.00	16.00	Score is neutral	27%
-Learning Difficulty	94.00	160.00	0.00	Lower is better	59%
Acquisition Efficiency	0.7700	1.00	0.10	Higher is better	77%
Transfer Problem	139.76	1699.00	0.00	Lower is better	8%
-Residual Performance Deficit	31.00	160.00	16.00	Lower is better	19%
-Residual Learning Difficulty	71.00	160.00	0.00	Lower is better	44%
-Additional Deficits	2.20	99.00	0.00	Lower is better	2%
--Physical Similarities	98.89	100.00	1.00	Higher is better	99%
--Functional Similarities	96.67	100.00	1.00	Higher is better	97%
Transfer Efficiency	0.9192	1.00	0.10	Higher is better	92%

With respect to transfer, the overall Transfer Problem percentage score (8%) indicated the rater thought a Soldier can be trained with the UCOFT and probably make a smooth transition to the gunner station on the Bradley Fighting Vehicle. The lower the score for this dimension the better. If the scores had been higher, that would indicate the training technology is either not very realistic or it has negative learning, such as controls that are critical but not replicated, or functions that differ from the real world situation. The relatively low Residual Performance Deficit subscore signified the rater did not perceive a substantial deficiency in Soldier performance on the subtasks or subtask skill components not trained with the UCOFT. However, the Residual Learning Difficulty percentage subscore (44%) indicated Soldiers would still need training and practice on these subtasks with the actual equipment to achieve the desired level of proficiency.

The very high percentage scores on Physical Similarities and Functional Similarities reflect the great attention devoted to designing the UCOFT to replicate the gunner's operational requirements in the Bradley Fighting Vehicle. Lastly, the Transfer Efficiency rating was in the desired direction (92% score); a high score is desired.

Table 11 presents the results from the same rater on the UCOFT using the Level II analysis of the Training Technology Evaluation Tool. The rater completed Level II prior to Level III. A comparison of the two sets of ratings using the percentage scores showed some differences on the Acquisition dimensions, particularly Performance Deficit and Learning Difficulty, which constitute the Training Problem dimension. The rater was asked about these differences in the post-evaluation interview. His comment regarding the Performance Deficit

dimension was that Level II only asks one question without regard to the subtasks, whereas Level III asks a similar question about each individual subtask, allowing him to respond with a more precise answer. He did not believe that his responses for Level II and Level III were that different. The smaller differences in the Level II and Level III percentage scores on the Learning Difficulty and Acquisition Efficiency dimensions were also attributed to differences in the questions and the specificity required.

Table 11
Gunnery Simulator Results, Level II Analysis

Score Comparisons, Level II					
Name of Training Technology **Gunnery Simulator**					
		16	# of Subtasks		
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	49.49	3590.00	0.41	Lower is better	1%
-Acquisition	41.21	1000.00	0.01	Lower is better	4%
-Transfer	8.28	2590.00	0.40	Lower is better	0%
Training Problem	35.39	100.00	0.01	Lower is better	35%
-Performance Deficit	75.00	100.00	1.00	Score is neutral	75%
-Learning Difficulty	755.00	1600.00	16.00	Lower is better	47%
Acquisition Efficiency	0.86	1.00	0.10	Higher is better	86%
Transfer Problem	8.00	259.00	0.40	Lower is better	3%
-Residual Performance Deficit	16.00	64.00	16.00	Lower is better	25%
-Residual Learning Difficulty	320.00	1600.00	16.00	Lower is better	20%
-Additional Deficits	0.00	99.00	0.00	Lower is better	0%
--Physical Similarities	1150.00	1600.00	16.00	Higher is better	72%
--Functional Similarities	1215.00	1600.00	16.00	Higher is better	76%
Transfer Efficiency	0.9661	1.00	0.10	Higher is better	97%

The other areas of discrepancy were the Residual Learning Difficulty, and the Physical and Functional Similarities scales. For each of these dimensions, the content of the scales changed substantially from the Level II to the Level III analysis. In addition, for Level II, the Residual Learning difficulty scale was a 100-point scale for each task, but for Level III there were six questions for each task with scales that ranged from 0 to 3 points. Consequently, with Level III there was less chance for variability in the ratings.

The questions on Physical and Functional Similarity in the Level II analysis are on individual subtasks, and how realistic the training technology is with regard to the subtasks. In Level III, the questions deal with the physical and functional similarity of each control and display to the operational equipment. The SME who rated the UCOFT felt the training focus in the Level III analysis was entirely different from that in Level II. The variation in the Level II and III questions appears to account for the differences in gunnery simulator ratings on these two dimensions.

IMI (Map Reading) – Levels I, II, and III Evaluation Results

Rater 2 evaluated a map reading IMI lesson on a CD-ROM using Levels I, II, and III. The eight subtasks evaluated are shown in Table 12. The overall task was "read a map." The target audience for this particular task was Soldiers who had no previous training in map reading. The IMI was designed to provide Soldiers with a minimum level of training so they could use a Soldier-based digital command and control system and the digital map display contained in the system's software. Conducting all analyses for this particular application provided an opportunity to compare the scores from all three levels using the same rater, and to provide insights regarding the impact of the differences in the questions at the three levels of analysis.

Table 12
Map Reading Subtasks

Subtask 1	Identify the topographic symbols on a map
Subtask 2	Identify the colors on a map
Subtask 3	Determine a direction
Subtask 4	Determine an azimuth
Subtask 5	Use scales
Subtask 6	Locate an object on a map using the Military Grid Reference System
Subtask 7	Determine a distance
Subtask 8	Identify terrain features

Tables 13 through 15 show the results from the Level I, II and III analyses. The percentage scores from the three analyses on the three summary dimensions, Total Effectiveness, Acquisition, and Transfer, were very similar. For the other component scores, in general, the Levels I and II results were very similar except for one dimension. However, the results from the Level III analysis differed from the other two levels in several areas.

In examining the dimensions that compose the Acquisition dimension, there were two major differences and one minor one. The Training Problem percentage score and the Learning Difficulty percentage scores were 50% lower (the desired direction) for the Level III analysis as compared to the Levels I and II analyses. This difference could reflect the fact that the questions on Training Problem and Learning Difficulty were more similar to each other in the Level I and II analyses than they were to the Level III questions (refer to Table 1). However, possible explanations for why the Level III scores were lower (in the desired direction) are not immediately obvious. One possible explanation resides in the nature of the scales used for these two dimensions. The ranges of the scales for the Level III questions were restricted, ranging from 1 to 10, 0 to 1, and 0 to 3. In contrast, the scales for the Level I and II questions ranged from 1 to 100, allowing for more variability in the SME judgments. Similarly, the precision of the Level III questions did not allow additional dimensions to influence the ratings, where in Levels I and II the rater could easily used additional factors in making judgments regarding the magnitude of the Training Problem and the Learning Difficulty for the target population.

Table 13
IMI Map Reading Results, Level I Analysis

Score Comparisons, Level I					
Name of Training Technology **Map Reading IMI**					
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	59.60	3000.00	0.02	Lower is better	2%
-Acquisition	36.00	1000.00	0.01	Lower is better	4%
-Transfer	23.60	2000.00	0.01	Lower is better	1%
Training Problem	36.00	100.00	0.01	Lower is better	36%
-Performance Deficit	90.00	100.00	1.00	Score is neutral	90%
-Learning Difficulty	40.00	100.00	1.00	Lower is better	40%
Acquisition Efficiency	1.00	1.00	0.10	Higher is better	100%
Transfer Problem	23.00	200.00	0.01	Lower is better	12%
-Residual Performance Deficit	20.00	100.00	1.00	Lower is better	20%
-Residual Learning Difficulty	15.00	100.00	1.00	Lower is better	15%
-Additional Deficits	20.00	100.00	0.00	Lower is better	20%
--Physical Similarities	75.00	100.00	1.00	Higher is better	75%
--Functional Similarities	55.00	100.00	1.00	Higher is better	55%
Transfer Efficiency	0.9747	1.00	0.10	Higher is better	97%

Table 14
IMI Map Reading Results, Level II Analysis

Score Comparisons, Level II					
Name of Training Technology **Map Reading IMI**					
	8	# of Subtasks			
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	68.02	2790.00	0.21	Lower is better	2%
-Acquisition	39.38	1000.00	0.01	Lower is better	4%
-Transfer	28.65	1790.00	0.20	Lower is better	2%
Training Problem	39.39	100.00	0.01	Lower is better	39%
-Performance Deficit	90.00	100.00	1.00	Score is neutral	90%
-Learning Difficulty	350.00	800.00	8.00	Lower is better	44%
Acquisition Efficiency	1.00	1.00	0.10	Higher is better	100%
Transfer Problem	25.63	179.00	0.20	Lower is better	14%
-Residual Performance Deficit	14.00	32.00	8.00	Lower is better	44%
-Residual Learning Difficulty	120.00	800.00	8.00	Lower is better	15%
-Additional Deficits	20.00	99.00	0.00	Lower is better	20%
--Physical Similarities	600.00	800.00	8.00	Higher is better	75%
--Functional Similarities	440.00	800.00	8.00	Higher is better	55%
Transfer Efficiency	0.8944	1.00	0.10	Higher is better	89%

Table 15
IMI Map Reading Results, Level III Analysis

Score Comparisons, Level III					
Name of Training Technology **Map Reading IMI**					
			8	# of Subtasks	
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	245.29	16990.00	0.00	Lower is better	1%
-Acquisition	182.89	8000.00	0.00	Lower is better	2%
-Transfer	62.40	8990.00	0.00	Lower is better	1%
Training Problem	160.00	800.00	0.00	Lower is better	20%
-Performance Deficit	80.00	80.00	8.00	Score is neutral	100%
-Learning Difficulty	16.00	80.00	0.00	Lower is better	20%
Acquisition Efficiency	0.8748	1.00	0.10	Higher is better	87%
Transfer Problem	44.13	899.00	0.00	Lower is better	5%
-Residual Performance Deficit	23.00	80.00	8.00	Lower is better	29%
-Residual Learning Difficulty	15.00	80.00	0.00	Lower is better	19%
-Additional Deficits	1.00	99.00	0.00	Lower is better	1%
--Physical Similarities	2.00	100.00	1.00	Higher is better	2%
--Functional Similarities	1.00	100.00	1.00	Higher is better	1%
Transfer Efficiency	0.7017	1.00	0.10	Higher is better	71%

A smaller difference (13% points) occurred on the Acquisition Efficiency rating, where the Level III percentage score was somewhat lower (not in the desired direction) as compared to Levels I and II. The difference in Acquisition Efficiency scores between Levels I and II and Level III can be found in the number and type of questions. Level I has one general question that references four instructional features for the SME to consider when rating the technology: clarity of the training objective, feedback, practice, and whether or not Soldier performance is recorded to a learning management system. Only one score can be given for this four-part question. Level II has four questions that cover the same information as Level I, except each area is given a score. Thus these questions were quite comparable, and the Acquisition Efficiency percentage scores for these two levels were identical (each 100%). In Level III, there are 11 Acquisition Efficiency questions for each subtask. These questions cover three of the four dimensions presented in Levels I and II, plus eight other instructional features. Because of these additional instruction features, Level III could be expected to yield a score that differs from Levels I and II.

With regard to the Transfer dimension, the Level II analysis percentage score for the Residual Performance Deficit was higher than Levels I and III. The percentage score was 44% for Level II, 29% for Level III, and 20% for Level I. The Level II rating indicated a higher level of deficit for the tasks that the training technology addresses. A careful examination of the nature of the questions revealed that each level of analysis had a slightly different focus (refer to Table 3). How these differences may have accounted for the results is discussed in the next paragraph.

In summary, all three levels of analysis focus on Residual Performance Deficits (skills not completely covered by the technology), but in different ways. At Level I, there is only one question about Residual Performance Deficit, while Levels II and III have one question about every subtask, but the response options differ for the two levels of analysis. Level I asks what percentage of the required skills and knowledge are not covered by the training technology. For Level II, the questions focus on whether the subtask was trained with the training technology and the extent to which the SME feels the subtask will be difficult for Soldiers to learn on the actual equipment/situation. For Level III, there is a similar focus, but the SME can indicate the degree to which the Soldier can perform the subtasks independently (without supervision) in the real world situation as a consequence of being exposed to the training technology. Although these differences in focus at the three levels of analysis may appear slight, they apparently did affect the SME who rated the IMI on map reading.

For non-simulator technologies such as IMI courseware, the Level III the Physical and Functional Similarity questions are cancelled out, as the Level III analysis specifically addresses displays and controls. This accounts for the low scores and percentages shown in Table 13. On the other hand, Levels I and II address physical and functional similarity with regard to subtasks, and ratings are made. For the IMI on map reading that was assessed, these ratings were the same for both levels of analysis.

The final area of difference was on Transfer Efficiency. The Level III analysis results reflected less certainty than Levels I and II. Level I asks one question, Level II has three questions, and Level III has one transfer question for each subtask. As the levels progressed, the percentage scores became lower. The focus of the question(s) also changes from level to level. Level I asks about transfer as related to the use of educational principles in the technology. Level II asks about the similarity of the subtasks to the real world, the extent of practice the Soldier gets with the technology, and the degree of realism in the practice sessions. Level III asks whether the amount of practice a Soldier will receive on each subtask will produce a job entry level of skill. Given the variation in the focus of the questions, one would not necessarily expect the scores on Transfer Efficiency to be similar across the three levels of analysis. The reduction in the score at Level III appears to result from the fact that the rater felt there should be more practice built into the lesson. The inference is that the technology scored well, but there is room for improvement, which might be resolved if more practice sessions were added.

Individual Weapons Simulator – Levels I and II Evaluation Results

Military SMEs were not available for the evaluations of the individual weapons simulator. Instead, three raters who had observed training on the training technology for the Infantry Forces Research Unit of the Army Research Institute performed two evaluations each. Rater 5, a human behavior scientist, performed the Level I evaluations. Rater 6, a graduate student intern in the field of training and education performed the Level II evaluations. Rater 10, also a human behavioral scientist, conducted Level I and II evaluations. The raters were not marksmanship SMEs. However, all were intimately involved with and knowledgeable about training and Soldier performance on the weapons simulator.

At the time the observations were made of the individual weapons simulator, the technology itself was relatively new. The simulator included scenarios that allowed squad leaders to train a nine-man squad on fire distribution and control. It also trained them on basic rifle marksmanship skills. It did not train Soldiers on the close combat optic (CCO), the thermal weapon sight (TWS), nor the infrared aiming light, which is used with night vision goggles. During pre and post evaluation interviews with the raters, all agreed that tasks supporting these sights and device should be incorporated in the technology in order to more completely support marksmanship training.

The overall task evaluated was restricted to "engage targets with an M-4/M-16 rifle." It did not include squad-level marksmanship skills. The 20 subtasks under this task are shown in Table 16. Since a high percentage of subtasks (8 of 20, 40%) were not trained with the marksmanship simulator, in the Level II analysis Raters 6 and 10 only evaluated the 12 subtasks that the simulator did train. An asterisk (see Table 16) marks the subtasks not evaluated.

Table 16
Engage Targets with an M-4/M-16 Subtasks

Subtask 1	Perform a function check on an M16-/M4-series weapon.
Subtask 2	Load and unload an M16/M4 magazine.
Subtask 3	Load an M16-/M4-series weapon.
Subtask 4	Unload an M16-/M4-series weapon.
Subtask 5	Correct malfunction of an M16-/M4-series weapon.
Subtask 6	Conduct shot grouping exercise (live fire).
Subtask 7	Conduct 25-meter zeroing.
Subtask 8	Engage single targets with the M16-/M4-series weapon.
Subtask 9	Detect and engage single timed targets with the M16-/M4-series weapon.
Subtask 10	Detect and engage single and multiple timed targets with the M16/M4 series weapon.
Subtask 11	Detect and engage timed targets with the M16-/M4-series weapon. Practice fire.
Subtask 12	Detect and engage timed targets with the M16-/M4-series weapon. Record fire.
Subtask 13*	Conduct 25-meter zeroing with the M68 close combat optic (CCO).
Subtask 14*	Detect and engage timed targets with the M16-/M4-series weapon and CCO. Practice fire.
Subtask 15*	Detect and engage timed targets with the M16-/M4-series weapon and CCO. Record fire.
Subtask 16*	Conduct 25-meter zeroing with the AN/PAS-13.
Subtask 17*	Detect and engage timed targets with the M16-/M4-series weapon and TWS. Practice Fire.
Subtask 18*	Detect and engage timed targets with the M16-/M4-series weapon and TWS. Record Fire.
Subtask 19*	Conduct boresighting with IR aiming laser at ten meters.
Subtask 20 *	Detect and engage multiple timed targets with the M16-/M4-series weapon and aiming laser while viewing through night vision goggles.

Note. * indicates subtasks not rated.

Raters 5 and 10 conducted a Level I analysis. Rater 5 conducted the analysis on the basis of the overall task of engage targets in accordance with the Level I analysis procedures. However, for this analysis, Rater 10 identified seven subtasks that were used as the basis for answering the Level I questions (see first column in Table 7). This was done to enhance the consistency of rater judgments across all the analysis questions.

The evaluation of the individual weapons simulator was also the first time the Training Technology Evaluation Tool was tried, with Rater 5 being the first person to evaluate any training technology with the Tool. The results of the Level I evaluations are shown in Tables 17 and 18.

Table 17
Individual Weapons Simulator Results, Level I Analysis, Rater 5

Score Comparisons, Level I					
Name of Training Technology **Marksmanship Training Simulator**					
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	47.81	3000.00	0.02	Lower is better	2%
-Acquisition	35.86	1000.00	0.01	Lower is better	4%
-Transfer	11.95	2000.00	0.01	Lower is better	1%
Training Problem	30.00	100.00	0.01	Lower is better	30%
-Performance Deficit	50.00	100.00	1.00	Score is neutral	50%
-Learning Difficulty	60.00	100.00	1.00	Lower is better	60%
Acquisition Efficiency	0.8367	1.00	0.10	Higher is better	84%
Transfer Problem	10.00	200.00	0.01	Lower is better	5%
-Residual Performance Deficit	20.00	100.00	1.00	Lower is better	20%
-Residual Learning Difficulty	50.00	100.00	1.00	Lower is better	50%
-Additional Deficits	0.00	100.00	0.00	Lower is better	0%
--Physical Similarities	60.00	100.00	1.00	Higher is better	60%
--Functional Similarities	80.00	100.00	1.00	Higher is better	80%
Transfer Efficiency	0.8367	1.00	0.10	Higher is better	84%

For Rater 5, the Level I summary percentage scores were quite low (Total Effectiveness, Acquisition and Transfer), indicating that the rater believed the technology was effective. Examination of the component Acquisition scores indicated the rater thought the tasks would be difficult to learn due to the Soldiers' existing skills and knowledge (Performance Deficit), and that the subtasks were perceived as being above average in terms of difficulty (Learning Difficulty). However, the Acquisition Efficiency score was relatively high indicating that the rater perceived the training technology to be good.

The results for Rater 10's Level I analysis are in Table 18. A primary difference between the two raters was that Rater 10 perceived the Training Problem to be more severe than Rater 5. Similarly, Rater 10 perceived the Transfer as a slightly greater problem than Rater 5. However, both raters agreed on ratings for Acquisition Efficiency, and Transfer Efficiency. The

differences between raters on the Training Problem were attributed to differences in their marksmanship experience and knowledge.

Table 18
Individual Weapons Simulator Results, Level I Analysis, Rater 10

Score Comparisons, Level I					
Name of Training Technology **Marksmanship Training Simulator**					
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	101.18	3000.00	0.02	Lower is better	3%
-Acquisition	71.30	1000.00	0.01	Lower is better	7%
-Transfer	29.88	2000.00	0.01	Lower is better	1%
Training Problem	61.75	100.00	0.01	Lower is better	62%
-Performance Deficit	95.00	100.00	1.00	Score is neutral	95%
-Learning Difficulty	65.00	100.00	1.00	Lower is better	65%
Acquisition Efficiency	0.8660	1.000	0.10	Higher is better	87%
Transfer Problem	25.00	200.00	0.01.00	Lower is better	13%
-Residual Performance Deficit	25.00	100.00	1.00	Lower is better	25%
-Residual Learning Difficulty	60.00	100.00	1.00	Lower is better	60%
-Additional Deficits	10.00	100.00	0.00	Lower is better	10%
--Physical Similarities	85.00	100.00	1.00	Higher is better	85%
--Functional Similarities	75.00	100.00	1.00	Higher is better	75%
Transfer Efficiency	0.8367	1.0000	0.10	Higher is better	84%

A Level II analysis on the individual weapons simulator was conducted by two raters: Rater 10 and Rater 6. These results are in Tables 19 and 20, respectively. Although the three summary scores did not differ substantially, the other scores showed differences in the SME assessments. Rater 10 perceived both the Training Problem and Transfer Problem as more severe than Rater 6, particularly the Training Problem. On the other hand, Rater 10 gave the Weapons Simulator higher ratings on Acquisition Efficiency and Transfer Efficiency. Consequently, the overall Acquisition and Transfer ratings from the two raters did not differ substantially.

Rater 10's Level I and II assessments were consistent despite the slight change in subtasks used for the analyses and the differences in the Level I and II tools. Acquisition Efficiency and Transfer Efficiency were given high ratings, and the Training Problem was perceived as being substantial. On Transfer Problem subscores, the Residual Performance Deficit was rated higher (not the desired direction) with the Level II analysis where each subtask was rated.

Table 19
Individual Weapons Simulator Results, Level II Analysis, Rater 10

Score Comparisons, Level II					
Name of Training Technology **Marksmanship Training Simulator**					
			12	# of Subtasks	
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	118.70	3190.00	0.31	Lower is better	4%
-Acquisition	71.43	1000.00	0.01	Lower is better	7%
-Transfer	47.26	2190.00	0.30	Lower is better	2%
Training Problem	52.25	100.00	0.01	Lower is better	52%
-Performance Deficit	95.00	100.00	1.00	Score is neutral	95%
-Learning Difficulty	660.00	1200.00	12.00	Lower is better	55%
Acquisition Efficiency	0.7300	1.00	0.10	Higher is better	73%
Transfer Problem	40.93	219.00	0.30	Lower is better	19%
-Residual Performance Deficit	31.00	48.00	12.00	Lower is better	65%
-Residual Learning Difficulty	537.00	1200.00	12.00	Lower is better	45%
-Additional Deficits	6.25	99.00	0.00	Lower is better	6%
--Physical Similarities	800.00	1200.00	12.00	Higher is better	67%
--Functional Similarities	745.00	1200.00	12.00	Higher is better	62%
Transfer Efficiency	0.8660	1.00	0.10	Higher is better	87%

Table 20
Individual Weapons Simulator Results, Level II Analysis, Rater 6

Score Comparisons, Level II					
Name of Training Technology **Marksmanship Training Simulator**					
			12	# of Subtasks	
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	58.72	3190.00	0.31	Lower is better	2%
-Acquisition	32.23	1000.00	0.01	Lower is better	3%
-Transfer	26.49	2190.00	0.30	Lower is better	1%
Training Problem	20.06	100.00	0.01	Lower is better	20%
-Performance Deficit	45.00	100.00	1.00	Score is neutral	45%
-Learning Difficulty	535.00	1200.00	12.00	Lower is better	45%
Acquisition Efficiency	0.6200	1.00	0.10	Higher is better	62%
Transfer Problem	18.42	219.00	0.30	Lower is better	8%
-Residual Performance Deficit	17.00	48.00	12.00	Lower is better	36%
-Residual Learning Difficulty	520.00	1200.00	12.00	Lower is better	43%
-Additional Deficits	0.00	99.00	0.00	Lower is better	0%
--Physical Similarities	635.00	1200.00	12.00	Higher is better	53%
--Functional Similarities	800.00	1200.00	12.00	Higher is better	67%
Transfer Efficiency	0.6952	1.00	0.10	Higher is better	70%

Games for Training – Level III Evaluation Results

The Army (Erwin, 2000) is looking more and more to video gaming technology to help train Soldiers, especially in light of the increasing presence of digital technologies and communications. The Marine Corps has also used this technology. They have bought off-the-shelf games and then had the manufacturer modify the games to meet their needs and standards. Given the emergence of training gaming technologies, it was of interest to determine the results from application of the Tool to this type of technology and to identify any potential training issues.

The specific training game that was evaluated also provided an opportunity to determine if rater preference or bias could be reduced or eliminated by focusing on the questions and responses. The premise was that since the questions in Level III are very detailed and query many possible nuances of a training technology, this type of analysis might force raters to take a more objective look at the training technology. As stated in the Method section, an opportunity presented itself to test this premise when it was discovered that two retired infantry NCOs working on the same gaming project had diametrically opposing views of the game's merit. Rater 3, a retired infantry master sergeant, had mechanized Infantry experience and experience in using computer driven simulators such as the UCOFT and the Close Combat Tactical Trainer (CCTT). Rater 4, a retired infantry sergeant first class, had light Infantry experience, had served as a Ranger instructor, and had no experience with computer-driven simulators while on active duty. During a pre-evaluation interview Rater 3 stated he thought the game had merit for training military operations in urban terrain, and that he probably had a positive bias toward simulators and computer-based training. Rater 4 stated that he thought the game was a "waste of time" and that nothing could replace boots on the ground when it came to training infantrymen in tactical skills. Yet he stated that he saw merit in interactive multi-media to help Soldiers acquire knowledge, so he was open to the idea of using computer-based games.

The overall task for the game was a squad engaging enemy forces in an urban setting. The subtasks that the squad and squad leader had to accomplish are listed in Table 21. These subtasks are a mix of individual and collective skills. However, the squad members in the game were not Soldiers but artificial intelligent entities that responded to the squad leader's commands. So when the subtask states, "Squad moves using concealment of smoke" for example, the squad leader is actually being evaluated as to how he controls his squad.

The results from Rater 3 and Rater 4 are in Tables 22 and 23, respectively. The three summary percentage scores (Total Effectiveness, Acquisition and Transfer) differed between the two raters, despite the low percentages for these scores. Rater 3's scores reflected a greater training value from the gaming technology than Rater 4's scores. Rater 4's scores for the gaming technology were the most negative ratings obtained in all the applications of the Training Technology Evaluation Tool cited in this report. Thus it appeared that Rater 4's reservations about gaming technology for training these skills were reflected in the Level III ratings. The extent to which these predispositions might have been reduced through actual application of the Tool itself could not be quantified. However, it is important to note that Rater 4's views seemed to improve after using the Tool. Although his comments concerning gaming technology were extremely negative during the pre-evaluation interview, he expressed a more positive opinion

during the post-evaluation interview and admitted that the technology had some merit. Yet, the caveat made in the pre-evaluation interview that "Nothing will replace boots on the ground when it comes to training infantrymen" was repeated in the post-evaluation interview.

Table 21
Training Game Subtasks

Subtask 1	Squad leader (SL) plans using the troop leading procedure
Subtask 2	SL addresses actions on chance contact with the enemy
Subtask 3	SL sends Situation Report (SITREP)
Subtask 4	Squad moves only after defenders' fires have been suppressed or obscured
Subtask 5	Squad moves using concealment of smoke
Subtask 6	SL prepares squad members for tactical movement
Subtask 7	Squad moves in built-up area
Subtask 8	Squad consolidates and reorganizes, as necessary
Subtask 9	Squad secures enemy prisoners of war (EPW), as required
Subtask 10	Squad determines when casualties are treated, as necessary
Subtask 11	SL sends end of mission (EOM) SITREP
Subtask 12	Squad continues operations as directed

Table 22
Training Game Results Level III Analysis, Rater 3

Score Comparisons, Level III

Name of Training Technology Game

12	# of Subtasks

Dimension/Component	Actual Scores Received During Evaluation	Maximum Possible Score	Minimum Possible Score	Score Interpretation	% of Maximum Score
Total Effectiveness	303.80	24990.00	0.00	Lower is better	1%
-Acquisition	247.37	12000.00	0.00	Lower is better	2%
-Transfer	56.43	12990.00	0.00	Lower is better	0%
Training Problem	140.25	1200.00	0.00	Lower is better	12%
-Performance Deficit	33.00	120.00	12.00	Score is neutral	28%
-Learning Difficulty	51.00	120.00	0.00	Lower is better	43%
Acquisition Efficiency	0.5670	1.00	0.10	Higher is better	57%
Transfer Problem	40.00	1299.00	0.00	Lower is better	3%
-Residual Performance Deficit	12.00	120.00	12.00	Lower is better	10%
-Residual Learning Difficulty	39.00	120.00	0.00	Lower is better	33%
-Additional Deficits	1.00	99.00	0.00	Lower is better	1%
--Physical Similarities	2.00	100.00	1.00	Higher is better	2%
--Functional Similarities	1.00	100.00	1.00	Higher is better	1%
Transfer Efficiency	0.7089	1.00	0.10	Higher is better	71%

Table 23
Training Game Results, Level III Analysis, Rater 4

Score Comparisons, Level III					
Name of Training Technology Game					
		12	# of Subtasks		
Dimension/Component	**Actual Scores Received During Evaluation**	**Maximum Possible Score**	**Minimum Possible Score**	**Score Interpretation**	**% of Maximum Score**
Total Effectiveness	1143.94	24990.00	0.00	Lower is better	5%
-Acquisition	529.70	12000.00	0.00	Lower is better	4%
-Transfer	614.24	12990.00	0.00	Lower is better	5%
Training Problem	348.75	1200.00	0.00	Lower is better	29%
-Performance Deficit	45.00	120.00	12.00	Score is neutral	38%
-Learning Difficulty	93.00	120.00	0.00	Lower is better	78%
Acquisition Efficiency	0.6584	1.00	0.10	Higher is better	66%
Transfer Problem	436.50	1299.00	0.00	Lower is better	34%
-Residual Performance Deficit	78.00	120.00	12.00	Lower is better	65%
-Residual Learning Difficulty	67.00	120.00	0.00	Lower is better	56%
-Additional Deficits	1.00	99.00	0.00	Lower is better	1%
--Physical Similarities	2.00	100.00	1.00	Higher is better	2%
--Functional Similarities	1.00	100.00	1.00	Higher is better	1%
Transfer Efficiency	0.7106	1.00	0.10	Higher is better	71%

When comparing the percentages of the maximum possible score for each dimension, the two raters clearly differed on three dimensions: Learning Difficulty, Residual Performance Deficit, and Residual Learning Difficulty. In each case, Rater 4 rated the game as less effective than Rater 3. There was also a smaller difference in the same direction with regard to Performance Deficit.

With respect to the Training Problem dimension, differences in the perceived Performance Deficit between Rater 3 and Rater 4 were revealed during the post-evaluation interview with Rater 4. Rater 4 believed that Soldiers would not be computer literate. Rater 4's corresponding Performance Deficit score reflected not only the perceived amount and difficulty of the knowledge and skills required by the subtasks, but also a deficit in learning how to operate the computer to perform the subtasks with the game.

Rater 4's score on Learning Difficulty was high (not the desired direction). However, Rater 3's score was in the low mid-range, and was substantially different from Rater 4's scores. Rater 4 indicated that the Soldiers he considered when making this rating were mechanized Infantrymen. As these Soldiers rarely train in urban environments, the skills they would have to learn were perceived to present some difficulties.

The Residual Performance Deficit and Residual Learning Difficulty subscores for Rater 4 were much higher (not the desired direction) than those for Rater 3. Rater 4, whose Residual Performance Deficit and Residual Learning Difficulty scores were in the high mid-range, believed that although Soldiers would have experienced urban operations training with the game

they would still need to perform the tasks on the ground. The game would not help them practice tasks such as using fragmentary grenades, flash bang grenades, crossing urban danger areas, and entering a building and entering a room. These factors explain why the Residual subscores for Rater 4 were in the negative direction. Yet, Rater 4 believed that the game would help Soldiers understand the tasks better and allow a leader to make better use of training time with a unit in an urban operations field environment.

As stated previously, the Training Technology Evaluation Tool did not eliminate Rater 4's initial "bias" about games as a training tool, but the Tool may have helped reduce bias. Of more importance was what the Tool in conjunction with rater interviews revealed about the specific training technology and its likely effectiveness. The pre- and post-evaluation interviews helped illuminate the reasons behind the scores. The most interesting aspect of the interviews for this particular technology was the discovery that what was originally thought to be a bias was not a personal bias. Instead, the different reactions to the game were based on the raters' assessment of the Soldiers' extant knowledge and skills, and how these would influence the learning that could occur from use of the training game technology. In addition, Rater 4 felt the game did not train all aspects of critical subtasks adequately.

Discussion

Subtask Selection

Both the gunnery simulator and the individual weapons simulator applications raised the question of how to proceed with the analysis when the training technology does not or cannot train all the relevant subtasks. This is an important issue to consider because SMEs must rate each subtask for some questions in the Level II and III analyses. The basic question is whether the analysis should reflect how well the training technology trains the subtasks for which it was designed or how well the training technology trains all the relevant subtasks that support the main task. One approach would be to conduct the analyses using both sets of subtasks and compare the two results. As indicated by the Level I sensitivity analyses, these results will differ as a function of subtask coverage.

Identifying each subtask per se is not required with the Level I analysis, as the SME does not have to rate each subtask. However, at this level the SME must determine the percentage of subtasks adequately covered by the training technology.

In Level III some questions, such as those on Acquisition Efficiency, assume the training technology covers each of the subtasks. Yet the SME cannot rate a subtask that is not covered. For example, the SME cannot answer the question on whether memory aids are used in the training technology for a subtask that is not trained. For Level III, the Training Technology Evaluation Tool should be modified to account for subtasks that are not trained.

If subtasks not covered by the technology are excluded at any level of analysis, then the evaluation reflects the adequacy of the training technology for the specified subtasks. However, training developers and trainers still must train these remaining subtasks using the operational equipment or other techniques. For example with the individual weapons trainer, the eight tasks

not trained on the technology are critical to Soldier proficiency. Soldiers will still need to be trained on them using operational equipment, ranges, and ammunition.

Another issue related to subtask selection is the relative scope, complexity, difficulty, or importance of each subtask. Again, this issue relates directly to the Level II and III analyses, as each subtask is often evaluated. Typically, the average of the subtask ratings is computed to provide an overall score, and each subtask is weighted the same. However, if the subtasks differ on critical performance dimensions, then the overall score for the training technology may not appropriately reflect the technology's training effectiveness. If, for example, many of the subtasks are small in scope and easy to perform, and there are only one or two more complex, difficult subtasks, then the small-scope, easy subtasks will contribute more to the final score than the few, yet more complex subtasks. The SME should not avoid rating all the appropriate subtasks, but attention should be paid to how the subtasks are conceptualized in the subtask list so inappropriate weighting does not occur.

Component Scores

Depending on the level of analysis, the three summary scores provided in the Scoring Guide are often too general and can obscure the relative strengths and weakness of the training technology being assessed. The assessments of the Training Technology Evaluation Tool indicated that examining the other component scores is essential. These other dimensions, such as Learning Difficulty, Acquisition Efficiency, and Residual Performance Deficit, provide the most important information on the relative strengths and weaknesses of the training technology being evaluated, such as whether the technology incorporates sound training principles, and whether there are major elements of the task that must still be trained in the operational environment or on the operational equipment. In addition, these subscores provide insights into why SMEs' ratings differ.

Level of Analysis

The choice of the level of analysis for a given training technology depends on what is known about the technology as well as the nature of the tasks/subtasks being assessed. No problems were encountered in applying the Level I analysis to the diversity of training technologies examined. However, it is advisable that the SME list the tasks being considered when conducting a Level I analysis, to enhance the consistency of the ratings across the component scores. This also enables the SME to make more precise estimates when asked such questions as "What percentage of tasks must the Soldier learn in the real world because they are not taught or practiced with the technology?" Lastly, a list of the subtasks considered helps evaluators interpret the results and clarify differences in SME ratings.

It should be noted that all the training technologies to which the evaluation Tool was applied existed; none were hypothetical or technologies in the planning stage. Thus, issues that might be encountered with the Level I analysis when there are many unknowns about the task to be trained or the proposed technology were not identified.

The Level II analysis also worked well. Compared to Level I, more information must be known about the tasks/subtasks to be trained, how the training technology functions, and the learning principles that are incorporated.

The SMEs encountered some problems in attempting to apply the Level III analysis to subtasks associated with the individual weapons simulator. The problem related to some of the Acquisition Efficiency questions, as they did not apply to some subtasks, and the SME could not indicate that the question did not apply. The SME could not indicate that the question did not apply, so the subtask was rated anyway. To illustrate this problem, when a task is very narrow in scope, it is difficult to answer questions regarding whether the training technology progresses from easy to hard aspects of the subtask, or whether examples are provided of all conditions under which the subtask is performed. On the other hand, if the subtask is broader in scope, then such questions can be answered. In addition, some of the questions under Acquisition Efficiency applied to training a group of subtasks with the technology rather than to each subtask, but the Training Technology Evaluation Tool did not allow the SME to answer from this broader perspective. Other options should be provided to the SME regarding the appropriate response in such instances, or the initial set of subtasks should be examined for their scope and complexity to ensure greater compatibility with the questions in the Tool.

Lastly, it was not always possible to identify the reasons for a substantial or unexpected change in scores when progressing from one level of analysis to the next. Score changes occurred only when the questions changed, but score changes were not automatic with question changes. In some cases, changes in scores reflected greater precision in the ratings. In other cases, they appeared to reflect changes in the aspect of training being assessed. A methodology to determine which factor was operative when going from one level of analysis to another was not within the scope of this research.

SME Ratings

In general, the results showed that SMEs tended to differ even when using the same level of analysis, but that a given SME was relatively consistent across the levels of analysis. These SME differences were attributed primarily to their experience with the tasks and subtasks being rated and their experience with Soldiers performing the tasks. This impacted their ratings of task and subtask difficulty, and their perception of the adequacy of the training standard incorporated in the training technology.

SME differences can be reduced by clearly defining the target population the SME must consider when applying the Training Technology Evaluation Tool to a particular training technology. To enhance the internal consistency of a Level I analysis, it is also recommended that the SME list the subtasks considered in the evaluation.

Interviewing the SMEs after they applied the Tool provided valuable insights into their assessments, insights that would not have been learned otherwise. In addition, interviews yielded information on what SMEs feel is important to train a Soldier in the domain of interest. It is recommended that this procedure be followed when using the Tool.

Refinements to the Tool

The rating scales in the Tool are not standardized. Some questions ask for a numerical percentage. Some questions are dichotomous, some are 10-point scales, and some are 100-point scales. In addition, most only have anchor point definitions for the two extreme points on the scale. Asking raters to distinguish among 100 different points without intermediate anchors is cognitively quite difficult and likely to introduce measurement error. Standardizing all or most of the scales on a common metric with definitions of scale points would reduce these problems, increase inter-rater reliability, and make it easier to compare the results from the different levels of analysis.

In addition, for some Tool dimensions, a high score is good; for other dimensions, a high score is not good. This inconsistency can easily lead to confusion regarding interpretation of the results, despite the Scoring Guide that indicates whether high scores are desired or not desired. It would be easier for users of the Tool to interpret the results if the formulae for all dimensions resulted in high scores as being desired. In addition, the maximum possible score for each dimension should be the same, facilitating direct comparisons across dimensions and across levels of analysis.

Application

It was apparent that it is inappropriate to compare training technologies in different task domains. For example, it would be inappropriate to try to compare how well Simulator X trains anti-armor tasks to how well Simulator Y trains indirect fire tasks. However, comparisons of training technologies that apply to the same task can be made. Under this condition, however, there will be greater consistency in the ratings if the same SME assesses each training technology of interest.

Each assessment required about thirty minutes to two hours to conduct, depending on the level of the evaluation and the number of subtasks being evaluated. Due to the automated aspects of the evaluation, a score was rendered as soon as the rater completed the forms.

The Scoring Guide provides a means to determine a particular training technology's strengths and weaknesses. This allows decision-makers to make better estimates of how a training technology fares (or will fare) in terms of specific areas such as educational principles, task difficulty, and the size of the residual learning deficit after training on the technology.

Summary

The assessments of the Training Technology Evaluation Tool showed that it had some face validity. For example, when applied to a well-established gunnery simulator, the high scores on the physical and functional similarities with the Level III analysis clearly reflected the attention that had been devoted to replicating the actual equipment in the gunnery simulator. In addition, differences between SMEs on their assessments of task difficulty and of the extent to which a given training technology approached the training in a real-world situation were also reflected in the Tool's output.

However, the assessments identified some problems with the Tool which were beyond the scope of the research to resolve. In particular, modifications should be made to the Level III analysis questions so that subtasks that are not trained can be included and scored appropriately. Also, in Level III when a question does not apply to a particular subtask, an appropriate response option should be provided, and the scoring algorithms modified accordingly. For all the analysis levels, it would be helpful if the maximum and minimum values were identical for all component scores. Scores should be normed for an average task, rather than dependent on the number of tasks. This would facilitate comparisons across levels of analysis. The "percentage of maximum score" created for the Scoring Guide in this research effort worked well for some scores, but not well for all scores in each level of analysis.

Whenever possible, more than one SME should rate the technology, and more than one level of analysis should be applied. Also, when subtasks within a task domain are not covered by the training technology being evaluated, two approaches are recommended. One is to evaluate the training technology on only the subtasks it trains (whether it trains all aspects of the task or only some aspects of the subtask). Then evaluate it with regard to all the subtasks within the task domain. These evaluations are best done with either the Level I or Level II analysis tools. A comparison of both results will provide the user a better picture of the strengths and weaknesses of the technology.

In summary, the three levels of the Training Technology Evaluation Tool can assist SMEs and training developers in evaluating training technologies during various stages of design and development. All three levels can be applied against a variety of training technologies, ranging from simulations and simulators, to games, and to interactive multimedia instruction.

The Training Technology Evaluation Tool should be used by an expert who understands the training principles addressed in the Tool and also the skills of the Soldiers in the target population. The Level I analysis assists such SMEs and training developers in understanding the extent to which the proposed design of an emerging training technology reflects many basic training principles. This analysis could help improve the technology's design, the training principles it incorporates, and the physical and functional similarities of the technology to the operational equipment/real world situation. Level II can be used to help evaluate the next level of readiness in the design and development cycle of training technologies. This analysis can be applied during beta testing and the initial fielding of newly emerging training technologies, which have little history or utilization. Level III allows trainers to take a harder look at the contributions made by current and legacy training technologies. This level of analysis can help identify which technologies are the most effective, to ascertain the less effective technologies, and to identify needed improvements to a training technology. Finally, the Tool can be used by trainers in the field to assist them in isolating the strengths and weaknesses any training technology might have, so that they can take advantage of the technology's strengths and overcome its weaknesses.

References

Department of the Army. (2003). *Rifle Marksmanship, M16A1, M16A2/3, M16A4 and M4 Carbine* (FM 3-22.9). Washington, D.C.: Headquarters, Department of the Army.

Department of the Army. (1994). *Mission training plan for the infantry rifle platoon and squad* (ARTEP 7-8-MTP). Washington, D.C.: Headquarters, Department of the Army.

Erwin, S. I. (2000). Video games gaining clout as military training tools. [Electronic version] National Defense Magazine, Retrieved December 22, 2004, from http://www.nationaldefensemagazine.org/issues/2000/Nov/Video_Games.htm

Narva, M. A. (1979). *Development of a systematic methodology for the application of judgmental data to the assessment of training device concepts* (ARI Research Memorandum 79-7). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A077 980)

Rose, A. M., Wheaton, G. R., & Yates, L. G. (1985a). *Forecasting device effectiveness: Volume I. Issues (*ARI Technical Report 680). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A159 576)

Rose, A. M., Wheaton, G. R., & Yates, L. G. (1985b). *Forecasting device effectiveness: Volume II. Procedures (*ARI Research Report 85-25). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A159 955)

Rose, A.M., Martin, A.W., & Yates, L.G. (1985). *Forecasting device effectiveness: Volume III. Analytic assessment of device effectiveness forecasting technique* (ARI Research Report 681). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A160 029)

Sticha, P. J., Gibbons, S., & Singer, M. J. (1993). *Development of a concept formulation process aid for analyzing training requirements and developing training devices* (ARI Research Report 1637). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Tufano, D.R., & Evans, R. A. (1982). *The prediction of training device effectiveness: A review of Army models* (ARI Technical Report 613). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A146 937)

Wheaton, G. R., Fingerman, P. W., Rose, A. M., & Leonard, R. (1976). *Evaluation of the effectiveness of training devices: Elaboration and application of the predictive model* (ARI Research Memorandum 76-16). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD A076 818)

Appendix A

Acronyms

A	Acquisition
AD	Additional deficits
AE	Acquisition efficiency
ARM	Advanced rifle marksmanship
BRM	Basic rifle marksmanship
CCO	Close combat optic device
CCTT	Close Combat Tactical Trainer
COBOL	Common business operating language
D	Learning difficulty
DEFT	Device Effectiveness Forecasting Technique
EOM	End of mission
EPW	Enemy prisoners of war
FM	Field Manual
FS	Functional similarity
FSA	Functional Similarity Analysis
IMI	Interactive multimedia instruction.
PDA	Personal digital assistant
PS	Physical similarity
PSA	Physical Similarity Analysis
RLD	Residual learning difficulty
RPD	Residual performance deficit
SITREP	Situation report
SL	Squad leader
SME	Subject matter expert
T	Transfer
TADSS	Training Aids, Devices, Simulators, and Simulations
TP	Training Problem
TRAINVICE	Training Effectiveness Device
TRP	Transfer problem
TT	Transfer efficiency
TWS	Thermal weapons sight
UCOFT	Unit Conduct of Fire Trainer
Σ	Total Effectiveness

Appendix B

A Sample of the Initial Training Assessment Worksheet

ASSESSMENT WORKSHEET					
What is the training deficiency?					
How does the technology resolve the training deficiency?					
FEASIBILITY:					
What is the Technical Readiness Level (TRL) of the technology?					
Is it matured, or is it concept (what is the road ahead, if concept)?					
What is the initial cost? Sustainment Cost? Power (Batteries/Generator)? Maintenance?					
What are the savings in Soldier/leader hours? Flight hours? OPTEMPO Mileage? Maintenance (Repair Parts)?					
Explain (if applicable) road ahead for negative responses!					
DOCTRINE:					
Does it require changes to Doctrine?	YES	NO	UNK	N/A	
ORGANIZATION:	YES	NO	UNK	N/A	
Does it require changes to Organizational structure?					
TRAINING:					On a scale of 1 – 5, rate the following: 1=UNSATISFACTORY, 2= POOR, 3=GOOD, 4=FAIR, 5=EXCELLENT.
Is this training realistic?	YES	NO	UNK	N/A	
Does it increase/decrease current safety requirements?	YES	NO	UNK	N/A	
Does it develop command of the basics?	YES	NO	UNK	N/A	
Does it build skill mastery?	YES	NO	UNK	N/A	
Does It replace current training hours?	YES	NO	UNK	N/A	
Does it focus on combat tasks?	YES	NO	UNK	N/A	
Does it teach any improper lesson?	YES	NO	UNK	N/A	
Does it support Mission Rehearsal Exercises (MRE)?	YES	NO	UNK	N/A	
					On a scale of 1 – 5, rate the following: 1-UNSATISFACTORY, 2-POOR, 3-GOOD, 4-FAIR, -EXCELLENT
Does it support unit Combat Readiness assessment?	YES	NO	UNK	N/A	
Does it provide task visualization?	YES	NO	UNK	N/A	
Does it adapt to different learning rates?	YES	NO	UNK	N/A	
Does it permit task repetition?	YES	NO	UNK	N/A	

Does it permit task repetition?	YES	NO	UNK	N/A	
Is it capable of degraded mode operations?	YES	NO	UNK	N/A	
Does it train motor skills?	YES	NO	UNK	N/A	
Does it increase retention of motor skills?	YES	NO	UNK	N/A	
Does it train cognitive skills?	YES	NO	UNK	N/A	
Does it increase retention of cognitive skills?	YES	NO	UNK	N/A	
Does it train psychomotor skills?	YES	NO	UNK	N/A	
Does it increase retention of psychomotor skills?	YES	NO	UNK	N/A	
Does it allow for increased degree of difficulty?	YES	NO	UNK	N/A	
Does it train individual tasks (whole/partial)?	YES	NO	UNK	N/A	
Does it train collective tasks (whole/partial)?	YES	NO	UNK	N/A	
Does it train the critical performance measures of the task?	YES	NO	UNK	N/A	
Does it provide immediate, understandable feedback?	YES	NO	UNK	N/A	
Does it facilitate the AAR process?	YES	NO	UNK	N/A	
Does it allow human facilitator assessment?	YES	NO	UNK	N/A	
Does it allow artificial assistant assessment?	YES	NO	UNK	N/A	
Does it allow analog data collection?	YES	NO	UNK	N/A	
Does it support live fire, tactical engagement exercises?	YES	NO	UNK	N/A	
Does it support OFW training requirements?	YES	NO	UNK	N/A	

					On a scale of 1 - 5, rate the following: 1=UNSATISFACTORY, 2=POOR, 3=GOOD, 4=FAIR, 5=EXCELLENT.
Does it support Interim Force training requirements?	YES	NO	UNK	N/A	
Does it support Legacy Force training requirements?	YES	NO	UNK	N/A	
Is it multi-functional? (Applicable to training and combat)	YES	NO	UNK	N/A	
Does it increase Instructor Certification requirements? (Hours)	YES	NO	UNK	N/A	
Does it reduce Instructor workload (tasks/skills)? (Hours)	YES	NO	UNK	N/A	
Does it require limited time to learn to use tool to train others?	YES	NO	UNK	N/A	
Does it permit easy access by trainer/Soldier being trained?	YES	NO	UNK	N/A	
MATERIEL					
Does it require modification of current equipment?	YES	NO	UNK	N/A	
Is it transportable/deployable?	YES	NO	UNK	N/A	
Will it operate at temperatures above 86 degrees Fahrenheit?	YES	NO	UNK	N/A	
Will it operate at temperatures below 32 degrees Fahrenheit?	YES	NO	UNK	N/A	
Will it operate in sand /dust/dirt/ salt/jungle/urban environment?	YES	NO	UNK	N/A	
Will it operate in/under water (how long, what depth), rain?	YES	NO	UNK	N/A	
Is increased performance worth weight and bulk to Soldier?	YES	NO	UNK	N/A	
Will Soldier accept the form and fit of the equipment?	YES	NO	UNK	N/A	
Will Soldier willingly use this device?	YES	NO	UNK	N/A	
Does it have greater than 90%OR Rate?	YES	NO	UNK	N/A	
Will it adapt to emerging technology (open architecture)?	YES	NO	UNK	N/A	
Is it computer based?	YES	NO	UNK	N/A	
Is it virtual simulation?	YES	NO	UNK	N/A	

					On a scale of 1 - 5, rate the following: 1=UNSATISFACTORY, 2=POOR, 3=GOOD, 4=FAIR, 5=EXCELLENT.
LEADER DEVELOPMENT:					
Does it require change in OES, NCOES, and WOES? (Hours)	YES	NO	UNK	N/A	
Does it support distance learning?	YES	NO	UNK	N/A	
Does it support distributed learning?	YES	NO	UNK	N/A	
PERSONNEL:					
Does it require changes to accession testing?	YES	NO	UNK	N/A	
Does it require changes to initial entry training?	YES	NO	UNK	N/A	
Does it require changes to Reenlistment program?	YES	NO	UNK	N/A	
FACILITIES:					
Does it require changes to buildings (Arms Rooms, Supply Rooms, Motor Pools, etc.)?	YES	NO	UNK	N/A	
Does it require modification/changes to ranges/training areas?	YES	NO	UNK	N/A	
Is it environment friendly?	YES	NO	UNK	N/A	

Appendix C

A Sample of the Revised Training Assessment Worksheet

Training Assessment Worksheet Version 1

Programmatic Dimensions **Note: You can add comments to yes/no questions at the end of this instrument. Be sure to associate your comments with the category identifier and question number.**

Requirement (R)

Category	Question	Narrative	Quantity	Units
R	1	What is the training need?		
R	2	How does this technology resolve the training need?		
R	3	Are there other technologies that meet this training need?		
R	4	How well does the training technology cover the actual tasks to be trained?		

Cost (C)

Category	Question	Narrative	Quantity	Units
C	1	What is the initial cost of the technology?		Dollars
C	1A	What facilities are required?		Feet
	1B	Does this technology require modification of any existing facilities? If yes, answer 1C. If not, skip to question 1D.	YES	NO
	1C	What is the cost of facility modifications?		Dollars
C	1D	What support personnel are required?		Personnel
C	1E	What, if any, specialized training is required for support personnel? List:____________________ ____________________ ____________________ ____________________ ____________________S how Manhours		Manhours
C	1F	What power requirements exist?		Volts/Cycles/Amps
C	2	What are the cost per setup in terms of manhours?		Manhours
C	3	What are the costs per iteration?		Dollars
C	3A	What are the costs in manhours for support personnel per iteration?		Manhours
C	3B	What are the power costs per iteration?		Dollars
C	3C	How many soldiers can be trained per iteration?		Number

Environmental Operations (EO)						
Category	Question	Narrative	Quantity	Units		
EO	1	What is the minimum operating temperature?		Degrees Farenheit		
EO	2	What is the maximum operating temperature?		Degrees Farenheit		
EO	3	What is the minimum operating relative humidity?		Percent		
EO	4	What is the maximum operating relative humidity?		Percent		
EO	5	Can the technology be operated in dusty conditions?	YES	NO		
EO	6	Can the technology be operated in the rain?	YES	NO		
EO	7	Is the technology transportable?	YES	NO		
EO	8	Is the technology deployable?	YES	NO		
EO	9	Will it operate at 24VDC?	YES	NO		
EO	10	Will it operate at 110VAC, 60Hz?	YES	NO		
EO	11	Will it operate at 220VAC, 50Hz?	YES	NO		
EO	12	What amperage is required?		Amps		
Environmental Storage (ES)						
Category	Question	Narrative	Quantity	Units		
ES	1	What is the minimum storage temperature?		Degrees Farenheit		
ES	2	What is the maximum storage temperature?		Degrees Farenheit		
ES	3	What storage limitations are there for humidity, if any?		Percent		
ES	4	Can the technology be stored in open weather?	YES	NO		
Safety (S)						
Category	Question	Narrative	Quantity	Units		
S	1	Are there any additional safety restrictions for support personnel?	YES	NO		
S	2	Are there any additional safety restrictions for soldiers to be trained?	YES	NO		
Accessibility (AC)						
Category	Question	Narrative	Quantity	Units		
AC	1	Does the technology allow for ease of access?	YES	NO		
AC	1A	Is there a scheduling requirement?	YES	NO		
AC	1B	How far in advance must scheduling be performed?		Days		
AC	2	Can this technology be freely accessed?	YES	NO		
AC	3	Does this technology allow for repetitive iterations?	YES	NO		

Learning and Training Dimensions						
Training Objectives (TO)						
TO	1	Are there training objectives associated with this training technology?	YES	NO		
TO	2	Are the objectives clearly stated and attainable?	YES	NO		
TO	3	Does this training technology allow the user to attain basic skill levels?	YES	NO		
TO	4	Does this training technology measure attainment of skill levels?	YES	NO		
TO	5	Does this technology allow for mastery of the learning objectives?	YES	NO		
TO	6	Is the training technology self-paced?	YES	NO		
TO	7	Are there skills required to complete the task that are not trained using this technology? Please list them._____	YES	NO		
TO	8	Are there any procedures or skills required to operate the technology that result in negative training (unrealistic or counter to proper procedures[for example a flight simulator designed to train 3 minute turns - but has no rudder input, or a marksmanship trainer that does not require sight manipulation to attain battle sight zero]]? Please list them._____	YES	NO		

Learning Domains (LD)						
LD	1	Does this technology fall in the cognitive domain?	YES	NO	NOTE 1	
LD	2	Does this technology fall in the psychomotor domain?	YES	NO	NOTE 2	
LD	3	Does this technology fall into both domains?	YES	NO	NOTE 3	
		NOTE 1 - if the answer is YES skip to the CD section	YES	NO		
		NOTE 2 - if the answer is YES skip to the PM section	YES	NO		
		NOTE 3 - if the answer is YES complete the CD and PM sections	YES	NO		
Cognitive Domain (CD)						
CD	1	Does this technology target any one or more of the following areas?	YES	NO		
CD	1A	**Knowledge** (Examples: Recite a policy. Quote information from memory. Knows the rules for a procedure.)	YES			
CD	1B	**Comprehension** (Examples: Rewrites the principles of war. Explain in one's own words the steps for performing a complex task. Translates an equation into a computer spreadsheet.)	YES	NO		
CD	1C	**Application** (Examples: Use a manual to calculate road march time. Apply the steps of the military decision-making process to a situation.)	YES	NO		
CD	1D	**Analysis** (Examples: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences. Troubleshoot a piece of equipment by using logical deduction. Recognize logical fallacies in reasoning. Gathers information from a department and selects the required tasks for training.)	YES	NO		
CD	1E	**Synthesis** (Examples: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure. Write a company operations order. Integrate training from several sources to solve a problem. Revises and process to improve the outcome.)	YES	NO		

Appendix D

The Training Technology Evaluation Tool Level I

The Level I Training Technology Evaluation Tool does not consider the number of subtasks to be trained. It is simple and will help training technology designers ensure that the technologies and training they are developing are effective and efficient. This appendix contains the questions that are present on each of the worksheets. They are organized in the order that they are presented in the Level I tool. Each question in this appendix also shows the means by which it is evaluated, and what the scale means.

Level I Training Problem

Performance Deficit

Question 1. Examine the statement of the training objective. Considering what you know about the typical Soldier's background, work experience, and prior training, what percentage of the skills and Knowledge required to meet the training objective will the Soldier still need to learn in order to reach criterion proficiency in the training technology?

Rating: Scale, 1-100

Scale Guidance:

1= None; the Soldier can already meet the training objective.

100 = All; the Soldier has to learn all of the skills and Knowledge needed to meet training objective.

Learning Difficulty

Question 2. Rate the difficulty the typical Soldier will have in acquiring the skills and Knowledge needed to meet the training objective.

Rating: Scale, 1-100

Scale Guidance:

1 = Very easy to learn; it will take practically no training or practice on the technology to learn the skills and Knowledges needed to meet the training objective.

100= Very difficult to learn; it will take a lot of training or practice on the technology to learn the skills and knowledges needed to meet the training objective.

Level I Acquisition Efficiency

Question 1. Examine information about the instructional features of the training technology, the training principles it incorporates, the program for its implementation, and the larger training context in which the technology is embedded. Consider the performance deficits you have identified in the previous worksheet and how use of the technology will over come these deficits.

To provide excellent training, the training technology should:

1. Make performance requirements of the training objective clearly understood to the Soldier.

2. Provide meaningful and understandable feedback to the Soldier regarding the results of his/her performance as soon as possible following his/her performance.
3. Provide sufficient practice where specific and hard-to-learn skills are involved.
4. Provide a record of the Soldier's performance.

Rating: Scale, 1-100
Scale Guidance:
1 = Poor training: the technology embodies few if any sound training principles and instructional features.
100 = Excellent training: the technology makes maximum use of sound training principles and instructional features.

Level I Transfer Problem

Residual Performance Deficit

Question 1. What proportion of the enabling skills and knowledges required in order to reach criterion proficiency on the operational equipment will the Soldier still have to learn? In other words, if there are skills and knowledges the training technology does not teach or practice, and they must be taught/trained in the real world situation, what percentage of all skills and knowledges do these represent?

Rating: Scale, 1-100
Scale Guidance:
1 = None; the Soldier can already meet the operational performance objectives.
100 = All; the Soldier has to learn all of the skills and knowledges needed to meet the operational performance objective(s).

Residual Learning Difficulty

Question 2. Consider the skills and knowledges that a Soldier who has been trained on the training technology must still acquire to perform at criterion level(s) on the operational equipment. Rate the difficulty of acquiring the remaining enabling skills and knowledges that are not taught or practiced using the training technology.

Rating: Scale, 1-100
Scale Guidance:
1 = Very easy to learn; it will take practically no training or practice on the operational equipment to learn the skills and knowledges needed to meet the operation performance objective.
100 = Very difficult to learn; it will take a lot of training or practice on the operational equipment to learn the skills and knowledges needed to meet the operational performance objective.

Physical Similarities

Question 3. Physical similarity is based on the similarity between physical characteristics of the training technology and those of the operational situation. The assessment is based on the

physical similarity (such as location, appearance, and feel) of displays, controls, and ambient conditions in the training and operational setting. Rate the physical similarity between the training technology and the operational equipment.

Rating: Scale, 1-100

Scale Guidance:

1 = Totally dissimilar; there would be a large difference, quite apparent to the Soldier and a large performance decrement, if the Soldier could perform at all. Specific instruction and practice would be required on the operational equipment to overcome the deficit.

100 = Identical; the Soldier would not notice a difference between the training technology and the operational equipment.

Functional Similarities

Question 4. Functional similarity is based on the Soldier's behavior in terms of the information flow from each display to the Soldier, and from the Soldier to each control. The assessment is made in terms of the amount of information transmitted from each display to each control and the type of information-processing activity performed by the Soldier. Rate the functional similarity between the training technology and operational equipment.

Rating: Scale, 1-100

Scale Guidance:

1 = Totally dissimilar; the Soldier acts on completely different types and amounts of information in the training technology and the operational equipment; the Soldier carries out different information-processing activities.

100 = Identical; the Soldier acts on the same types and amounts of information in the training technology and the operational equipment; the Soldier carries out the same information-processing activities.

Transfer Efficiency

Question 1. Consider the training objective and descriptions of the operational and training technology situations when answering this question. Consider the instructional features and training principles that are included in the training technology to increase the probability that the skills and knowledges acquired on the technology will be used effectively in the real world situation. Rate how well the training technology will promote transfer to the real world situation.

Rating: Scale, 1-100

Scale Guidance:

1 = Poor transfer; the technology embodies few, if any, sound training principles and instructional features to promote transfer to the operational situation.

100 = Excellent transfer; the technology makes maximum use of sound training principles and instructional features to promote transfer to the real world situation.

Appendix E

The Training Technology Evaluation Tool Level II

The Level II Training Technology Evaluation Tool considers the number of subtasks to be trained. It is slightly more complex and will help evaluate training technologies that are in the prototype stage of development. This appendix contains the questions that are present on each of the worksheets. They are organized in the order that they are presented in the Level II tool. Each question in this appendix also shows the means by which it is evaluated, and what the scale means.

Level II Training Problem

Performance Deficit

Question 1. Examine the statement of the training objective. Considering what you know about the typical Soldier's background, work experience, and prior training, what percentage of the skills and knowledges required to meet the training objective will the Soldier have to learn in order to reach criterion proficiency in the training technology?

Rating: Scale, 1-100
Scale Guidance:
1= None; the Soldier can already meet the training objective.
100 = All; the Soldier has to learn all of the skills and knowledges needed to meet training objective.

Learning Difficulty

Question 2. Rate the difficulty the typical Soldier will have in learning to perform each subtask on the training technology.

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Very easy to learn; it will take practically no training or practice on the technology to reach criterion proficiency on this subtask.
100= Very difficult to learn; it will take a lot of training or practice on the technology to reach criterion proficiency on this subtask.

Level II Acquisition Efficiency

Question 1. For what percentage of the subtasks that must be learned does the training technology make the criterion performance requirements explicit to the Soldier?

Rating: Scale, 1-100
Scale Guidance:
1 = None: performance requirements are not made explicit to the Soldier.
100 = All: performance requirements are made explicit to the Soldier.

Question 2. For what percentage of the subtasks that must be learned does the training technology provide practice?

Rating: Scale, 1-100

Scale Guidance:

1 = None: Practice is not provided on any of the subtasks which must be learned.

100 = All: Practice is provided on all the subtasks which must be learned.

Question 3. For what percentage of the subtasks that must be learned does the training technology provide qualitative feedback to Soldiers.

Rating: Scale, 1-100

Scale Guidance: 1 = None: Practice is not provided on any of the subtasks which must be learned.

100 = All: Practice is provided on all the subtasks which must be learned.

Question 4. For what percentage of the subtasks that must be learned does the training technology provide a record of the Soldier's performance.

Rating: Scale, 1-100

Scale Guidance: 1= None: records of Soldiers' performance are not provided for any of the subtasks which must be learned.

100 = All: records of performance are provided for all the subtasks which must be learned.

Level II Transfer Problem

Residual Performance Deficit

Question 1. Question: Assume that the Soldier can perform all the subtasks comprising the training objective on the training technology. For each subtask associated with the operational performance objective enter a value from 1 to 4 based on the definitions indicated below.

Rating: See Scale Guidance. .

Scale Guidance:

1 = This subtask was represented in the training objective: most Soldiers will be able to perform this subtask with a minimal exposure to or practice on the operational equipment (real world).

2 = This subtask was not represented in the training objective; but most Soldiers will be able to perform this subtask with minimal exposure to or practice on the operational equipment.

3 = This subtask was represented in the training objective: but most Soldiers will not be able to perform this subtask with minimal exposure to or practice on the operational equipment.

4 = This subtask was not represented in the training objective: most Soldiers will not be able to perform this subtask with minimal exposure to or practice on the operational equipment.

Residual Learning Difficulty

Question 2. Rate the difficulty the typical Soldier will have in learning to perform each subtask on the operational equipment/situation, after being trained with the training technology.

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance:

1 = Very easy to learn; it will take practically no training or practice on the operational equipment to reach criterion proficiency on this subtask.

100 = Very difficult to learn; it will take a lot of training or practice on the operational equipment to reach criterion proficiency on this subtask.

Physical Similarities

Question 3. Physical similarity is based on similarity between physical characteristics of the training technology and those of the operational situation. The assessment is based on the physical similarity (locations, appearance, feel of displays, controls, etc) and ambient conditions in the operational and training subtasks. Rate the physical similarity between each operational subtask and its counterpart, if any, in the training technology.

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance:

1 = Totally dissimilar; although the subtask is represented in the training technology, there would be a large difference, quite apparent to the Soldier, and a large performance decrement, if the Soldier could perform the subtask at all. Specific instruction and practice would be required for this subtask on operational equipment to overcome the deficit.

100 = Identical; the Soldier would not notice a difference between the training technology and the operational equipment for this subtask.

Functional Similarities

Question 4. Functional similarity is based on the Soldier's behavior in terms of the information flow from each display to the Soldier and from the Soldier to each control. The assessment is made in terms of the amount of information transmitted from each display to each control and the type of information-processing activity performed by the Soldier. Rate the functional similarity between each operational subtask and its counterpart, if any, in the training technology.

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance:

1 = Totally dissimilar; for this subtask the Soldier acts on completely different amounts and types of information in the training technology and the operational equipment; the Soldier carries out different information-processing activities.

100 = Identical; for this subtask the Soldier acts on the same amounts and types of information in the training technology and the operational equipment; the Soldier to be trained carries out the same information-processing activities.

Transfer Efficiency

Question 1. What percentage of the subtasks that must be learned in the technology are realistic and relevant in the sense that they are similar to the subtasks that are performed in the real world?

Rating: Scale, 1-100

Scale Guidance:

1 = None; the subtasks are not realistic, relevant, or similar to those in the real world.

100 = All; the subtasks are realistic, relevant, and similar to those in the real world.

Question 2. For what percentage of the subtasks that must be learned in the technology are conditions of practice late in the training approximate those found in the real world?

Rating: Scale, 1-100

Scale Guidance:

1 = None; late in training the conditions of practice do not approximate those likely to be encountered in the real world.

100 = All; late in training the conditions of practice are made to approximate those in the real world on all of the subtasks the Soldier must learn in the technology.

Question 3. For what percentage of the subtasks that must be learned in the technology is an extensive amount of practice given?

Rating: Scale, 1-100

Scale Guidance:

1 = None; not even a single task is practiced extensively.

100 = All; every subtask that a Soldier must learn in the technology is practiced extensively.

Appendix F

The Training Technology Evaluation Tool Level III

The Level III Training Technology Evaluation Tool is the most complex and detailed of the three levels of evaluation. It is designed to evaluate legacy and newly emerging training technologies. The rater must know much about the technology to conduct a Level III analysis. This appendix contains the questions that are present on each of the worksheets. They are organized in the order that they are presented in the Level III tool. Each question in this appendix also shows the means by which it is evaluated, and what the scale means.

Level III Training Problem

Performance Deficit

Question 1. Examine the description of the subtasks or tasks that comprise the training objective. Considering what you know about the typical Soldier's background, work experience, and prior training, level of proficiency, rate the current level of proficiency on each subtask.

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance: Enter a number from 10 to 1 using the following definitions:

10 = No experience, training, or familiarity with this subtask. Cannot perform this subtask.

9 = Has only limited knowledge about this subtask. Cannot be expected to perform the subtask. Has had orientation only.

7 = Has received a complete briefing on the subtask. Can perform the subtask only if assisted during every step. Requires much more training and experience. Has had familiarization training only.

4 = Understands the subtask to be performed. Can perform the subtask in the technology. Needs more practice under supervision. Has had procedural training.

1 = Has complete understanding of the subtask. Can do the subtask completely and accurately without supervision. Has received skill training.

Learning Difficulty

Questions 1-6.

Question 1. Are job or memory aids intended to be used in performing the subtask on the training technology?

Rating: See Scale Guidance. Each subtask is evaluated.

Scale Guidance:

If job aids are used enter "0"

If job aids are not used enter "1"

Question 2. How many steps are required to do the subtask?

Rating: See Scale Guidance. Each subtask is evaluated.

Scale Guidance:

If the subtask contains <10 steps enter "0"
If the subtask contains >10 steps enter "1"

Question 3. Is there a requirement to perform the steps in the subtask in a definite sequence?
Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If all or most of the steps in a subtask must be performed in a specified order, enter "1"
If the order in which steps are performed are not critical, then enter "0"

Question 4. Does the subtask have a natural logic so that the Soldier knows when he/she is doing it correctly? *Note*: Some tasks have a logical or natural sequence, like fixing a tire. Others have seemingly arbitrary steps such as troubleshooting tasks. Some contain illogical and arbitrary steps.
Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If the subtask contains logical and natural flow, enter "0"
If the subtask does not, enter "1"

Question 5. What are the mental or thinking requirements of the subtask? *Note*: Repetitive physical subtasks require little to no thinking. Many subtasks that look easy have very complex mental requirements, such as planning an attack. Consider the number of internal decisions or calculations that must be made to accomplish the subtask. Also consider if job aids are necessary.
Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If there are few mental requirements to accomplish the task enter "0"
If the subtask is mentally demanding, enter "3"

Question 6. What are the motor control demands of the subtask? *Note*: Motor control refers to coordinated precision finger and hand movements such as tracking a target.
Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If motor control demands are small, enter a "0"
If motor control demands are great, enter a "3"

Level III Acquisition Efficiency

Questions 1-11. Examine information about the instructional features of the training technology, the training principles it incorporates, the program for its implementation, and the larger training context in which the technology is embedded. Consider the deficits you should have identified during normal training. Rate how well use of the technology will overcome these deficits on each subtask.

Question 1. To what extent does the training technology make the training objective clear to the Soldier, and clarify the Soldier's proficiency with regard to that objective?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all: Neither the training objective nor the Soldier's standing relative to the objective is made explicit.
100 = Completely: the training objective and the Soldier's standing relative to that objective are made explicit throughout training.

Question 2. To what extent does the Soldier begin with an easy component of the subtask and progress to more difficult components in the training technology?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance: 1 = Not at all: the material to be learned is not sequenced in terms of learning difficulty.
100 = Completely: the material to be learned is sequenced in terms of learning difficulty from easy to hard; the sequence of instruction can be tailored to individual Soldier capabilities.

Question 3. To what extent does the training technology provide Soldiers with knowledge of results of their performance and positive reinforcement (feedback)?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; the training technology provides Soldiers with no feedback about their performance.
100 = Completely; the training technology provides Soldiers with explicit feedback about the adequacy of their performance.

Question 4. To what extent does the training technology provide for repetition/practice of the material to be learned?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Minimum practice; there is no practice, rehearsal, or repetition of subtasks.
100 = Maximum practice; there is extensive practice, repetition, rehearsal of subtasks; the amount of practice is tailored to the individual Soldier.

Question 5. To what extent does the training technology provide for prompting/cueing early in training and gradually fade as training progresses?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; prompts and cues are not used.
100 = Completely; the training technology provides prompting and cueing early in training and gradually fades them out late in training.

Question 6. To what extent does the training technology chunk material to be learned into small blocks or steps appropriate to the complexity of the subtasks and the capabilities of the training technology?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; the material to be learned is not organized into chunks.
100 = Completely; the material to be learned is organized into chunks; the chunks are tailored to the complexity of the material and the capabilities of the Soldiers.

Question 7. To what extent does the training technology use memory aids when these are practicable and available?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; although memory aids could be used, the training system does not employ them.
100 = Completely; the training system employs memory aids to facilitate learning.

Question 8. To what extent does the training technology start with a wide tolerance band for correct performance, narrowing the band as training progresses?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; error tolerances are not varied as training progresses; there is no shaping of behavior.
100 = Completely; error boundaries are broad at the beginning of training and become narrow late in training.

Question 9. To what extent does the training technology provide examples of all of the conditions under which the subtask will be performed in the training environment?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; the technology presents only one version of each subtask.
100 - Completely; the technology represents a range of conditions under which the subtasks are to be performed.

Question 10. To what extent does the training technology present the Soldier with different scenarios or situations that would require similar responses - in other words - the same task is performed under different circumstances?

Rating: Scale, 1-100. Each subtask is evaluated.
Scale Guidance:
1 = Not at all; different situations requiring the same or similar responses are not presented.
100 = Completely; different situations requiring the same or similar responses are presented.

Question 11. To what extent can the training system manipulate subtask difficulty, knowledge of results, reinforcement, practice, prompting, blocks of material, tolerance bands and stimulus and response conditions as a function of the performance of the Soldier?

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance:
1 = Not at all; the technology provides a lock-step instruction; training is not adaptive.
100 = Completely; the technology provides for adaptive training; the program of instruction is varied as a function of Soldier performance.

Level III Transfer Problem

Residual Performance Deficit

Question 1. Consider descriptions of the subtask(s) that comprise the training objective, the subtask(s) that comprise the operational performance objective, as well as the descriptions of the training technology and operational equipment, including their controls and displays. Assume the typical Soldier can perform all of the subtask(s) comprising the training objective on the training technology (that is the Soldier has reached criterion proficiency on each subtask in the training technology).

For each subtask associated with the operational performance objective, enter a number from 10 to 1 using the following definitions:

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
10 - No experience, training, or familiarity with this subtask. Cannot perform this subtask on the operational equipment.
9 - Has only limited knowledge about this subtask. Cannot be expected to perform the subtask. Has had an orientation only.
7 - Has received a complete briefing on the subtask. Can perform the subtask only if assisted in every step. Requires much more training and experience. Has had familiarization training.
4 - Understands the subtask to be performed. Can perform the subtask in the trainer. Needs more practice under supervision. Has had procedural training.
1 - Has a complete understanding of the subtask. Can do the subtask completely and accurately without supervision. as received skill training.

Residual Learning Difficulty

Questions 2-7. For each subtask, answer the following six questions regarding the Soldier's performance on the operational equipment.

Question 2. Are job or memory aids intended to be used in performing the subtask on the operational equipment?

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If the subtask is taught or tested with the use of memory aids, enter "0"
If the subtask is taught or tested without the use of memory aids, enter "1"

Question 3. How many steps are required to do the subtask on the operational equipment?

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:

If the subtask contains less than 10 steps, enter "0"
If the subtask contains 10 steps or more, enter "1"

Question 4. Is there a requirement to perform the steps in the subtask in a definite sequence on the operational equipment?

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If all or most of the steps must be performed in a specified order, enter "1"
If the order in which the steps are performed is not critical, enter "0"

Question 5. Does the subtask have a built-in logic so that the Soldier knows when he/she is performing it correctly on the operational equipment?

Definition - Some subtasks consist of steps that form a logical or natural sequence, like fixing a tire or changing a light bulb. Others have steps that appear arbitrary, like troubleshooting subtasks. Some contain a mix of "natural" and "unnatural" steps, such as safety steps that often break a natural flow and logic of a subtask.

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If the subtask contains a built in logic, enter "0"
If the subtask does not contain a natural flow, enter "1"

Question 6. What are the mental or thinking requirements of the subtask on the operational equipment?

Definition - Repetitive, physical subtasks that require almost no mental work (loader on a tank). Many subtasks that look easy have very complex mental requirements, such as planning an attack or troubleshooting a complex piece of equipment. Consider the number of internal decisions or calculations that must be made in choosing your answer. Also consider the impact of any job aid.

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If there are few mental requirements, enter "0"
If the subtask is mentally challenging or complex, enter "3"

Question 7. What are the motor control demands of the subtask on the operational equipment?

Definition - Motor control refers to precise finger, hand, or arm movements, not large body movement. Sheer physical strength is not a factor. For example, lifting a weight or changing a tire does not require much motor control.

Rating: See Scale Guidance. Each subtask is evaluated.
Scale Guidance:
If the motor control demands are small, enter "0"
If the motor control demands are great, enter "3"

Physical Similarities

Instructions for Questions 8 and 9. Questions 8 and 9 are primarily related to simulators of equipment or tactical situations. If the training technology is a simulator then answer Question

8 as it pertains to the physical similarities between the operational situation and the simulator, and answer Question 9 as it pertains to the functional similarities between the operational equipment and the simulator.
(In other words, for question 8 if the simulator's controls and displays are an exact replica of the operational equipment, you would score 100 for Question 8. If there is some difference then you would lower your score. For question 9, if the functions of the controls and displays in the simulation exactly duplicate the functions of the controls and displays of the operational equipment, you would score 100.)

NOTE: If you are evaluating a training technology that is NOT a simulator, then you would score a "4" in Question 8" and score Question 9 as a "2". DO NOT USE any other numbers. In either case, be sure to list the number of controls and the number of displays you evaluated. For non-simulator training technologies, enter 1 for each.

Question 8.

Question: Physical similarity is based on the similarity between physical characteristics of the training technology and those of the operational situation. The assessment is based on the physical similarity such as location, appearance, and feel of displays, controls, and ambient conditions in the operational and training tasks.
Rate the physical similarity between each operational display and control and its counterpart, if any, in the operational equipment.

Rating: Scale, 1-100. Each display/control is evaluated.
Scale Guidance:
1 = Totally dissimilar; although the display/control is represented in the training technology, the Soldier would notice a large difference, and there would be a large performance decrement on the operational equipment, if the Soldier could perform the subtask at all. Specific instruction and practice would be required for this display/control.
100 = Identical; the Soldier would not notice a difference between the training technology and the operational situation for this display/control.

Functional Similarities

Question 9. Functional similarity is based on the Soldier's behavior in terms of the information flow from each display to the Soldier and from the Soldier to each control. The assessment is made in terms of the amount of information transmitted from each display to each control and the type of information-processing activity performed by the operator.
Rate the functional similarity between each operational display and control and its counterpart, if any, in the operational equipment.

Rating: Scale, 1-100. Each control/display is evaluated.
Scale Guidance:
1 = Totally dissimilar; for this display/control the Soldier acts on completely different information in the training technology and the operational situation; the Soldier carries out different information-processing activities..

100 = Identical; for this display/control the Soldier acts on the same types of information in the training technology and the operational situation; the Soldier carries out the same information-processing activities.

Transfer Efficiency

Question 1. To what extent will the training technology permit Soldiers to practice in the technology until they can demonstrate a job entry skill on this subtask?

Rating: Scale, 1-100. Each subtask is evaluated.

Scale Guidance:

1 = Not at all; the amount of practice Soldiers receive will not enable them to perform at job entry skill levels.

100 = Completely; the amount of practice Soldiers receive will enable them to perform at a job entry skill level.

Appendix G

Formulae

Note.: The number of **i**'s preceding the formula name indicates that the formula relates to the Level, such as iPD, which is found in Level I. These are the base formulae names. Each Level is discussed in detail.

iPD = performance deficit – the Soldier's ability to perform the required task prior to being trained. Measured on a scale of 1 to 100, with 1 meaning the Soldier can already meet the training objective(s). The far extreme of the scale, a rating of 100, means the Soldier has to learn all of the skills and knowledge needed to meet training objective(s).

iD = difficulty inherent in overcoming the deficit – how hard will it be to train the Soldier to enable him to acquire the skills and knowledge required to meet the training objective(s) using the training technology. Measured on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the technology to learn the skills and knowledge needed to meet the training objective(s), and 100 meaning it will take a lot of training or practice on the technology to learn the skills and knowledge needed to meet the training objective(s).

Both iPD and iD are used to determine overall performance deficits.

iRating = how utilization of the technology will overcome the performance deficit – measures of the effectiveness of the training technology and its sound use of training principles. Measured on a scale of 1 to 100, with 1 signifying that the system embodies few if any sound training principles and instructional features. The opposite end of the scale means that the system makes maximum use of sound training principles and instructional features.

iRating is used to determine Acquisition Efficiency.

iRPD = Residual Performance Deficit – after the Soldier has been trained on the training technology, what must he still learn and practice to be proficient on the operational equipment or situation, using the operational equipment as the training vehicle. Measured on a scale of 1 to 100, with 1 meaning that the Soldier can already meet the operational performance objectives. No further training is required to achieve the training objectives on the operational equipment. The opposite end of the scale, the score of 100, represents the fact that the Soldier has to learn all of the skills and knowledge needed to meet the operational performance objective(s) using the operational equipment. A score of 100 would seem to almost indicate that training on the technology was a waste of time.

iRLD = Residual Learning Difficulty - is a rating of the residual learning difficulty to acquire the remaining enabling skills and knowledge – after the Soldier has been trained using the training technology, and assuming he has mastered all the skills and knowledge required to achieve the technology objective(s), how difficult would it be to train him on those tasks (subtasks,

knowledge, skills) that may only be trained on the operational equipment or situation. Measured on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the operational equipment to learn the skills and knowledge needed to meet the operational performance objective(s). The rating of 100 represents the fact that it will take a lot of training or practice on the operational equipment to learn the skills and knowledge needed to meet the operational performance objective(s).

iPS = physical similarity – what are the physical similarities between the training technology and the operational equipment or situation? Also rated on a scale of 1 to 100, with 1 representing there would be a large noticeable difference, quite apparent to the Soldier and a large performance decrement, given that the Soldier could perform at all; specific instruction and practice would be required on the operational equipment after transfer to overcome the deficit. The rating of 100 means that the two are identical and that the Soldier would not notice a difference between the training technology and the operational equipment at the time of transfer.

iFS = functional similarity – do the controls that exist within the training technology perform the same or similar functions in the operational equipment or situation?
Rated on a scale of 1 to 100 with 1 representing functions that are totally dissimilar and the Soldier must react to completely different types and amounts of information and carry out different information processing activities. A score of 100 indicates that functions are identical and the information and information processing is the same as it is on the operational equipment or situation.

iTTRting = transfer efficiency rating – considers the training objective(s) and descriptions of the operational and training technology situations and the instructional features and training principles that are included in the training technology to increase the probability that the skills and knowledge acquired on the technology will be used effectively in the operational situation. Rated on a scale of 1 to 100 with 1 representing poor transfer; the technology embodies few, if any, sound training principles and instructional features to promote transfer to the operational equipment or situation. A rating of 100 represents excellent transfer; the technology makes maximum use of sound training principles and instructional features to promote transfer to the operational equipment or situation.

iiiPS/CDP – physical similarity/control display physical similarity - In Level III, physical similarities takes on a new dimension. It now compares the physical similarities of the controls and displays between the training technology and the operational situation. The subject matter expert (SME) must list the displays and controls. For example, the driver's station for the Bradley Fighting Vehicle in the Close Combat Tactical Trainer (CCTT) has an accelerator pedal. So the description would indicate whether or not the accelerator pedal looks like the accelerator pedal in a real Bradley. Then the SME would evaluate the physical similarity of the accelerator pedal with an actual Bradley (not its function – only the physical similarity). This is scored on a scale of 1 to 100, with 1 indicating that the control or display is totally dissimilar to the actual control or display, and 100 indicating that the control or display is identical. The Bradley CCTT accelerator pedal would score 100, since it looks and is placed exactly like the one in the real Bradley. If it looked different but was in the same place, the SME could score it using a lesser

number. For example, a button on a computer keyboard that represented the accelerator pedal would score more than 0 but would remain low on the scale.

iiiFS/CDF – functional similarity/control display functional similarity. Functional similarities also take on a new and important dimension in Level III. Do the controls and displays act the same in the training technology as they do in the actual vehicle? The subject matter expert (SME) must list the displays and controls. Using the same analogy as in physical similarity, when the accelerator pedal is depressed the vehicle will display terrain passing at a faster speed and the speedometer will show that the vehicle is going faster. The description the SME would have to enter could be "Accelerator pedal manipulation indicates increase or decrease in speed relative to the terrain and the speedometer." This area is also scored on a scale of 1 to 100. One represents total dissimilarity in functions from the operational situation, and 100 represents identical functions to those found between the training technology and the operational situation.

NOTE: Since IMI or computer gaming can never fully represent the operational situation in order to make evaluations more simple, the instructions tell the SME to enter "4" under CDP and a "2" under CDF.

CCn – the number of controls found in the training technology – e.g., accelerator pedal, steering wheel, brake pedals, ignition, and transmission selector would all be controls found in the driver's station of a Bradley Fighting Vehicle at the CCTT. If there are none, the SME should enter "1" in the block to preclude a divide by zero error.

CDn – the number of displays found in the training technology – i.e. vision blocks with video input, instruments such as fuel gauge, speedometer, tachometer, etc. If there are none, the SME should enter "1" to preclude a divide by zero error.

Level I Specific Formulae

Training Problem (TP) is the product of the **Performance Deficit** (iPD) and **Learning Difficulty** (iD). **Performance Deficit** refers to the Soldier's ability to perform the required task prior to being trained. It is measured on a scale of 1 to 100, with 1 meaning the Soldier can already meet the training objective(s). The far extreme of the scale, a rating of 100, means the Soldier has to learn all of the skills and knowledge needed to meet training objective. **Learning Difficulty** (iD) is how hard will it be to train the Soldier on the training technology to enable him to acquire the skills and knowledge required to meet the training objective. It is measured on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the technology to learn the skills and knowledge needed to meet the training objective(s), and 100 meaning it will take a lot of training or practice on the technology to learn the skills and knowledge needed to meet the training objective. TP ranges from 0.01 to 100. PD and D ranges are 1 to 100.

Formula: TP = iPD * iD/100
Training Problem equals the Performance Deficit multiplied by the Learning Difficulty divided by 100.

Acquistion Efficiency (*AE*) is obtained directly from the rating of training deficiencies in the training technology. The **Rating** (iRating = how utilization of the technology will overcome the performance deficit) assesses the effectiveness of the training technology and its sound use of training principles. It is measured on a scale of 1 to 100, with 1 signifying that the technology embodies few if any sound training principles and instructional features. The opposite end of the scale means that the technology makes maximum use of sound training principles and instructional features. AE ranges from 0.1 to 1. iRating ranges from 1 to 100.

Formula: AE = SQRT (iRating/100)
Acquisition Efficiency equals the square root of the result of iRating divided by 100.

Acquisition (A) is the index reflecting the acquisition or training component. This is obtained by combining the training problem (i**TP**) and acquisition efficiency (i**AE**) indexes. *A* ranges between 0.10 and 100. An effective device has a relatively low A.

Formula: A = iTP/iAE
Acquisition equals the Training Problem divided by the Acquisition Efficiency.

Additional Deficits (AD) is derived by subtracting **Functional Similarity** (iFS) from **Physical Similarity** (iPS). Whenever FS>PS, AD equals 0. AD ranges from 1 to 100. Functional Similarity is the similarity in which controls function in the simulator and in the actual real life equipment. Physical Similarity is the visual representation of the actual equipment in the simulator and how close they resemble each other. Each similarity dimension is assessed on a scale from 1 to 100.

Formula: AD = iPS – iFS, unless FS > PS then AD = 0
Additional Deficits is the Performance Similarity minus the Functional Similarity.

Note. AD's value is set to the variable getAD. Where getAD equals PS-FS. The variable preAD is used to determine whether or not AD is positive or negative. If negative, AD is set to 0, otherwise it is the value of AD.

Transfer Problem (TRP) is the product of the **Residual Performance Deficit** (iRPD) and **Residual Learning Difficulty** (iRLD) plus the AD score. **Residual Performance Deficit** pertains to the situation after the Soldier has been trained on the training technology; specifically, what must he still learn by using the operational equipment or situation as the means of training. It is measured on a scale of 1 to 100, with 1 meaning that the Soldier can already meet the operational performance objective. No additional training on the operational equipment is required to achieve the training objective. The opposite end of the scale, a score of 100, represents the fact that the Soldier has to learn all of the skills and knowledge needed to meet the operational performance objective by training on/in the operational equipment or situation. A score of 100 would seem to indicate that training on the technology was a waste of time. **Residual Learning Difficulty** is a rating of the difficulty to acquire/learn the remaining enabling skills and knowledge after the Soldier has been trained with the training technology, and assuming he has mastered all the skills and knowledge required to achieve the technology objective. The question is how difficult would it be to train him on the skills and knowledge that

must be acquired by training on/in the operational equipment or situation. The rating is measured on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the operational equipment/situation to learn the skills and knowledge needed to meet the operational performance objective. The rating of 100 represents the fact that it will take a lot of training or practice on the operational equipment/situation to learn the skills and knowledge needed to meet the operational performance objective.

Formula: TRP = (iRPD *iRLD/100) + iAD
Transfer Problem equals the Residual Performance Deficit multiplied by the Residual Learning Difficulty divided by 100 then AD is added. It ranges from .01 to 200.

Transfer Efficiency (TT) is a computed index similar to AD. The **Rating** (iTTRating) considers the training objective and the instructional features and training principles that are included in the training technology. The rating is the extent to which the instructional and training features will increase the probability that the skills and knowledge acquired on the training technology will be used effectively in the operational situation. Transfer Efficiency is rated on a scale of 1 to 100 with 1 representing poor transfer; the technology embodies few, if any, sound training principles and instructional features to promote transfer to the operational equipment or situation. A rating of 100 represents excellent transfer; the technology makes maximum use of sound training principles and instructional features to promote transfer to the operational equipment or situation.

Formula: TT = SQRT (iTTRating/100)
Transfer Efficiency equals the square root of the Rating divided by 100. It ranges from 0.10 to 1.00.

Transfer (T) is an index that reflects the transfer component of device evaluation and is obtained by combining the **Transfer Problem** (TRP) and the **Training Efficiency** (TT) indexes. The index is the transfer analog of A (the index reflecting the acquisition or training component). T ranges between .1 and 200.

Formula: T = iTRP/iTT
Transfer equals the Transfer Problem divided by the Training Efficiency.

Total Effectiveness (Σ) is the index that represents overall training technology effectiveness. It is obtained by combining the **Acquisition** (A - the index reflecting the acquisition or training component) and the **Transfer** (T – the index that reflects the transfer component of the technology evaluation) indexes.

Formula: Σ= A + T
Total Effectiveness equals the Acquisition index added to the Transfer index.

Formulae for Level I analysis:

Σ = A + T

A = iTP/iAE

T = iTRP/iTT

TP = iPD * iD/100

AE = SQRT (iRating/100)

AD = iPS – iFS, unless FS > PS then AD = 0

TRP = (iRPD *iRLD/100) + iAD

TT = SQRT (iTTRating/100)

Level II Specific Formulae

Number of Subtasks = n

Training Problem (TP) is the **Performance Deficit** multiplied by the sum of the **Learning Difficulty** (iiD) ratings on each subtask and then divided by the results from multiplying the number of subtasks (n) by 100. Performance Deficit (iiPD) is the Soldier's ability to perform the required task prior to being trained. It is measured on a scale of 1 to 100, with 1 meaning the Soldier can already meet the training objective. The far extreme of the scale, a rating of 100, means the Soldier has to learn all of the skills and knowledge needed to meet the training objective. **Learning Difficulty** (iiD) is how hard it will be to train the Soldier to enable him to acquire the skills and knowledge required to meet the training objective when using the training technology. Each subtask is rated on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the technology to learn the needed skills and knowledge. One hundred (100) means it will take a lot of training or practice on the technology to learn the skills and knowledge needed to meet the training objective. iiTP ranges from 1 to 100. iiPD ranges from 1 to 100. The total number of subtasks evaluated determines the ranges for iiD.

Formula: iiD = $\left(\sum_{i=1}^{n} D_i\right)$

Formula: TP = iiPD * $\left(\sum_{i=1}^{n} D_i\right)$**/(100 *n)**

Acquistion Efficiency (AE) is obtained from the square root of the result of the sum of the four ratings of the training technology's application of training principles divided by 400. The **Rating** (iiRating = how utilization of the technology will overcome the performance deficit) measures the effectiveness of the training technology and its sound use of training principles as applied to each subtask. It is measured on a scale of 1 to 100, with 1 signifying that the technology embodies few if any sound training principles and instructional features. The opposite end of the scale means that the technology makes maximum use of sound training principles and instructional features. Four training principles are included in the evaluation. AE ranges from .1 to 1.

Formula: AE = SQRT ($\sum_{i=1}^{n}$ iiRating$_i$)/400 (NOTE: For AE, "n" will always equal 4.)

Acquisition (A) is the index reflecting the acquisition or training component. This is obtained by dividing the training problem (**TP**) by the acquisition efficiency (**AE**) index. *A* ranges are affected by the number of subtasks evaluated in Learning Difficulty (iiD) within **TP**. An effective device has a relatively low A.

Formula: A = iiTP/iiAE

Additional Deficit (AD) is the total of the pair-wise differences between Physical Similarity and Functional Similarity divided by the number of subtasks. **Functional Similarity** (iiFS) refers to the extent to which the controls that exist within the training technology perform the same or similar functions in the operational equipment or situation. It is rated on a scale of 1 to 100, with 1 representing functions that are totally dissimilar and the Soldier must react to completely different types and amounts of information and carry out different information processing activities. A score of 100 indicates that functions are identical and the information and information processing is the same as it is on the operational equipment or situation. **Physical Similarity** (iiPS) refers to the extent to which the physical similarities between the training technology and the operational equipment or situation are the same. It is also rated on a scale of 1 to 100, with 1 representing there would be a large difference, quite apparent to the Soldier, and a large performance decrement. Specific instruction and practice would be required on the operational equipment after transfer to overcome the deficit. The rating of 100 means that the two are identical and that the Soldier would not notice a difference between the training technology and the operational equipment. Whenever FS>PS, then (PS-FS) = 0. AD ranges from 0 to 99.

Formula: AD = $\sum_{i=1}^{n}$ (iiPS – iiFS)$_i$/n, when FS > PS then (PS-FS) = 0

Transfer Problem (TRP) is the sum of the **Residual Performance Deficit** (iiRPD) multiplied by the Residual Learning Difficulty (iiRLD) divided by 40, and the Additional Deficit (AD) is added to this result.

Residual Performance Deficit reflects the situation after the Soldier has been trained on the training technology. What must he still learn and practice to be proficient on the operational equipment or situation, using the operational equipment as the training vehicle? It is measured on a scale of 1 to 4 for each subtask. A score of 1 indicates that the subtask was represented in the training technology, and most Soldiers will be able to perform the subtask. A score of 2 indicates that the subtask was not represented in the training technology, but most Soldiers will still be able to perform the subtask. A score of 3 indicates that the subtask was represented in the training technology, but that most Soldiers will not be able to perform the subtask. A score of 4 indicates that the subtask was not represented in the training technology, and most Soldiers will not be able to perform the subtask.

The **Residual Learning Difficulty** (iiRLD) is a rating of the residual learning difficulty to acquire the remaining enabling skills and knowledge after the Soldier has been trained using the training technology, assuming he has mastered all the skills and knowledge required to achieve

the technology objective. How difficult would it be to train him on those tasks that may only be trained on the operational equipment or situation. It is measured on a scale of 1 to 100, with 1 meaning it will take practically no training or practice on the operational equipment to learn the skills and knowledge needed to meet the operational performance objective. The rating of 100 represents the fact that it will take a lot of training or practice on the operational equipment to learn the skills and knowledge needed to meet the operational performance objective.

Formula: $\mathbf{TRP = \sum_{i=1}^{n} [(iiRPD_i * iiRLD_i)/40] + iiAD}$

Since the number of subtasks affects RPD and RLD, a standard minimum and maximum range is dependent on the number of subtasks evaluated. For example if ten subtasks were evaluated, the minimum TRP score would be 0.25 and the maximum score would be 199. However, if the evaluated training technology had 15 subtasks, the minimum score would be 0.375 and the maximum score would be 249. A technology with 20 subtasks would have a minimum score of 0.5, and a maximum score of 299.

Transfer Efficiency (TT) is a computed index similar to AE. The **Rating** (iiTTRating) considers three areas: (1) the training objective, (2) descriptions of the operational and training technology situations, and (3) the instructional features and training principles that are included in the training technology. Each area is rated on a scale of 1 to 100 with 1 representing poor transfer; the technology embodies few, if any, sound training principles and instructional features to promote transfer to the operational equipment or situation. A rating of 100 represents excellent transfer; the technology makes maximum use of sound training principles and instructional features to promote transfer to the operational equipment or situation. TT ranges from 0.1 to 1.

Formula: $\mathbf{TT = SQRT [(\sum_{i=1}^{n} iiTTRating_i)/300]}$

Transfer (T) is an index that reflects the transfer component of training technology evaluation and is obtained by dividing the **Transfer Problem** (TRP) by the **Transfer Efficiency (TT)** indexes. Since the number of subtasks evaluated affects T, ranges cannot be developed except on a case-by-case basis. For example, if the training technology has ten subtasks to be evaluated, the minimum value for T would be 0.25, while the maximum would be 1990. For a training technology with 20 subtasks, the minimum score would be 0.5, while the maximum score would be 2990.

Formula: T = iiTRP/iiTT

Total Effectiveness (Σ) is the index that represents overall training technology effectiveness. It is obtained by adding the **Acquisition** (A - the index reflecting the acquisition or training component) and the **Transfer** (T – the index reflecting the transfer component of the technology evaluation) indexes.

Formula: $\Sigma = A + T$
Formulae for Level II Analyses:

$\Sigma = A + T$

A = iiTP/iiAE

T = iiTRP/iiTT

TP = iiPD *($\sum_{i=1}^{n}$ D_i)/(100 *n)

AE = SQRT ($\sum_{i=1}^{n}$ $iiRating_i$)/400 (NOTE: For AE, "n" will always equal 4.)

Formula: AD = $\sum_{i=1}^{n}$ $(iiPS - iiFS)_i$/n, when FS > PS for each comparison then (PS-FS) = 0

TRP = $\sum_{i=1}^{n}$ [($iiRPD_i$ *$iiRLD_i$)/40] + iiAD

TT = SQRT [($\sum_{i=1}^{n}$ $iiTTRating_i$)/300]

iiD = ($\sum_{i=1}^{n}$ D_i)

Level III Specific Formulae

Number of Subtasks = n

Training Problem (TP) is the **Performance Deficit** (PD_w) sum multiplied by the **Learning Difficulty** (D_{sum}) sum, and the product of these two sums is then divided by the number of subtasks.

Performance Deficit (PD_w) relates to the Soldier's ability to perform the required subtasks prior to being trained. The performance deficit can be assigned a rating of 10, 9, 7, 4, or 1 for each subtask that contributes to overall task proficiency. The number of subtasks affects the maximum and minimum range for PD_w. PD_w is the sum of these ratings across all subtasks.

Learning Difficulty (D_{sum}) reflects ratings for each subtask on six task characteristics, which impact learning difficulty. D_{sum} is the sum of the six task characteristic ratings across all subtasks. Each D_{sum} rating can be assigned a value of either 0 or 1 for Questions 1 through 4, and either 0 or 3 for Questions 5 and 6. The minimum score for D_{sum} will always be 0. The number of subtasks evaluated will affect the maximum score for $D_{sum.}$ and for TP.

Formula: TP = (PD_w * D_{sum})/n

Acquisition Efficiency (AE) is obtained from the ratings on the training efficiency of the training technology. The **Rating** (iiiRating = how utilization of the technology will overcome the performance deficit) measures the effectiveness of the training technology and its sound use of training principles. The technology is rated on the extent to which the training on each subtask incorporates 11 training principles. Each rating is made on a scale from 1 to 100, with 1 signifying that the technology does not embody the particular training principle or instructional feature. The opposite end of the scale (100) means that the technology makes maximum use of the particular training principle or instructional feature.

For each subtask, compute the average rating over the 11 training principles

$$R_i = \sum_{i=1}^{11} (Rating_i/11)$$

Acquisition Efficiency is computed by taking the square root of the sum of the subtask ratings divided by the product of 100 times the number of subtasks. AE ranges from 0.1 to 1.

Formula: $AE = SQRT\left(\sum_{i=1}^{n} R_i/(100*n)\right)$

Acquisition (A) is the index reflecting the acquisition or training component. This is obtained by dividing the Training Problem (iii**TP**) by the Acquisition Efficiency (iii**AE**). The number of subtasks evaluated affects the range. The minimum value will always be 0. The maximum value will be 1000.

Formula: A = iiiTP/iiiAE

Additional Deficit (AD) in Level III analysis is based on the operational equipment controls and displays (not tasks or subtasks). Additional deficit stems from differences in **Functional Similarity (FS)** and **Physical Similarity (PS)** between the operational world and the training technology.

Functional Similarity (iiiFS) addresses whether the controls that exist within the training technology perform the same or similar functions in the operational equipment. FS is rated on a scale of 1 to 100 for each control and display. One (1) represents functions that are totally dissimilar and the Soldier must react to completely different types and amounts of information and carry out different information processing activities. A score of 100 indicates that functions are identical and the information processing is the same as it is on the operational equipment.

Physical Similarity (iiiPS) addresses the physical similarities between the training technology and the operational equipment. Each control and display is rated on a scale of 1 to 100. One (1) represents a large difference in physical similarities, which would be quite apparent to the Soldier and creates a large performance deficit. Specific instruction and practice would be required on the operational equipment to overcome the deficit. The rating of 100 means that the two are identical and that the Soldier would not notice a difference between the training technology and the operational equipment.

CCn – the number of controls found in the training technology – i.e. accelerator pedal, steering wheel, brake pedal, ignition, and transmission selector would all be controls found in the driver's station of a Bradley Fighting Vehicle at the CCTT. If there are none, the rater should enter "1" in the block to preclude a division by zero error.

CDn – the number of displays found in the training technology – i.e. vision blocks with video input, instruments such as fuel gauge, speedometer, tachometer, etc. If there are none, the rater should enter "1" to preclude a division by zero error.

AD is the total of the pair-wise differences between Physical Similarity and Functional Similarity divided by the number of Controls and Displays. When calculating the difference between PS and FS, if FS is greater than PS, the result is converted to 0. Negative numbers are never used.

Formula: $AD = [\sum_{i=1}^{CCn+CDn} (iiiPS - iiiFS)_i] / (CCn + CDn)$, **when FS > PS for each comparison then (PS-FS) = 0**

Transfer Problem (TRP) is the residual performance and learning difficulty that a Soldier experiences after being trained using the training technology and attempting to perform those subtasks on the operational equipment. The **Residual Performance Deficit ($iiiRPD_w$)** is what the Soldier will not know how to do on the operational equipment after training to standard on the training technology. It is measured on a scale of 1 to 10, with 1 meaning that the Soldier can already meet the operational performance objectives. No further training is required to achieve the training objectives on the operational equipment. The opposite end of the scale, the score of 10, represents the fact that the Soldier has to learn all of the skills and knowledge needed to meet the operational performance objective using the operational equipment. The **Residual Learning Difficulty ($iiiRLD_{sum}$)** is a rating of the residual learning difficulty to acquire the remaining enabling skills and knowledge on the operational equipment after the Soldier has been trained using the training technology. There are four questions concerning factors that can make learning a subtask easier, and are rated as either a 1 or a 0. If the presence of the factor makes the subtask easier to learn it is rated as 0. If the factor is not present or makes the subtask more difficult to learn, it is rated as a 1. There are two questions that concern the mental and physical difficulty of the subtask and are rated as either a 3 or a 0. If a subtask is mentally or physically easy to learn, it is rated as a 0. If it is difficult, it is rated as a 3. **$iiiRLD_{sum}$** can range from 0 to 10.

TRP is the sum of the products of the Residual Performance Deficit and the Residual Learning Difficulty divided by the number of subtasks, and the Additional Deficit (AD) is added to this result. TRP can range from 0 to 199.

Formula: $TRP = [\sum_{i=1}^{n} (iiiRPD_w * iiiRLD_{sum})_i / n] + iiiAD$

Transfer Efficiency (TT) is a computed index similar to AE. The **Rating** (iiiTTRating) considers the extent to which practice on the training technology will result in the Soldier being proficient in the operational situation. It is rated on a scale of 1 to 100 with 1 representing poor transfer; where practice on the technology will not result in skill at job entry level. A rating of 100 represents excellent transfer. Each subtask is rated. TT ranges from 0.1 to 1.

Formula: $TT = SQRT (\sum_{i=1}^{n} iiiTTRating_i)/100)$

Transfer (T) is an index that reflects the transfer component of device evaluation and is obtained by dividing the **Transfer Problem (TRP)** by **Transfer Efficiency (TT)**. The minimum value for T will always be 0. The maximum value for T is 1990.

Formula: T = iiiTRP/iiiTT

Total Effectiveness (Σ) is the index that represents overall training technology effectiveness. It is obtained by adding the **Acquisition** (A) and the **Transfer** (T) indexes.

Formula: $\Sigma = A + T$

Formulae for Level III Analyses:

$$\Sigma = A + T$$

$$A = iiiTP/iiiAE$$

$$T = iiiTRP/iiiTT$$

$$TP = (PD_w \times D_{sum})/n$$

$$R_i = \sum_{i=1}^{11} (Rating_i/11)$$

$$AE = SQRT\left(\sum_{i=1}^{n} R_i/(100*n)\right)$$

$$AE = SQRT\left(\sum_{i=1}^{11} R_i/(100*11)\right)$$

Formula: $$AD = \left[\sum_{i=1}^{CC+CD} (iiiPS - iiiFS)_i\right] / (CCn + CDn)$$, **when FS > PS then (PS-FS) = 0**

$$TRP = \left[\sum_{i=1}^{n} (iiiRPD_w * iiiRLD_{sum})_i / n\right] + iiiAD$$

$$TT = SQRT\left(\sum_{i=1}^{n} iiiTTRating_i\right)/100)$$